DOCUMENT RESUME

ED 177 464 CS 004 706

AUTHOR	Anderson, Linda M.; And Others
TITLE	The First-Grade Reading Group Study: Technical Report of Experimental Effects and Process-Outcome Relationships. Volume I. R&D Report No. 4070.
INSTITUTION	Michigan State Univ., East Lansing.; Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY	Michigan State Univ., East Lansing. Inst. for Research on Teaching.; National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Oct 78
CONTRACT	4-0-76-0073; OB-NIE-B-78-0216
NOTE	339p.; For related documents, see CS 004 705 and CS 005 017-018; Volume II is only available from the Center
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	Grade 1; *Grouping (Instructional Purposes); Primary Education; Reading Instruction; Reading Programs; *Reading Research; *Small Group Instruction; *Teaching Models

ABSTRACT

This report is related to a study of first grade reading groups and presents the results of testing an instructional model developed from the integration of research and knowledge about how young children function in a classroom, especially within the small group format. The background and methodology of developing the instructional model is analyzed along with the achievement data for treatment effects. The results of group comparisons and regression analyses are discussed, the overall results of the study and a revision of the instructional model are summarized, and future experimental studies of teaching are suggested. The instructional model tested for first grade reading groups, a summary of the observation system used in the study, and a glossary of terms used to describe data from the study are appended. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED177464

The First-grade Reading Group Study: Technical Report of Experimental Effects and Process-Outcome Relationships

Volume I

Linda M. Anderson
Carolyn M. Evertson
R&D Center for Teacher Education
The University of Texas at Austin

and

Jere E. Brophy
Michigan State University

R&D Report No. 4070
(COET No. 78-6)

October, 1978

Research and Development Center for Teacher Education
The University of Texas at Austin

CS004706

The study was supported in part by the National Institute of Education, Contract OB-NIE-B-78-0216, The Research and Development Center for Teacher Education, The University of Texas at Austin. In addition, a portion of Dr. Brophy's time was supported by the National Institute of Education, Contract 4-0-76-0073, Institute for Research on Teaching, Michigan State University. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

Acknowledgements

Several persons made important contributions to the conduct of this study and the preparation of this report. The data were collected by Carol Greenhalgh, Dr. Jane Ogden, and Harry Selig. Data analyses were performed by Tom Linsley, Robert Morgan, Dr. Jeanne Martin, and Dr. Don Veldman. Valuable assistance was offered during the planning of the study by Dr. John Crawford and Linda Harris Rogers.

The manuscript was prepared by Susanne Kaboord, Betty Ancheta, Karen Kington, Suzanne Rohrer, Elijah Udom, and Debby Yarborough. Tables were organized and compiled by Mary Jane Leahy. In addition, many student research assistants performed onerous but vitally important tasks of organizing and checking data:

Julie Ballard
Berlye Bell
Mollie Fernandez
Jesús Gaspar
Luis Gonzalez
Teddie Melhart
Melvin Miller
Paul Persons
Kim Peterson
Norman Plotkin
Norton Rosenthal
Miriam Sparks
Ben Youngblood

Within the Austin Independent School District, several persons were instrumental in allowing the study to take place. Dr. Freda Holley, Coordinator of Research and Evaluation, and Dr. M. George Bowden, Director of Elementary Education, deserve special thanks, as do the principals of the nine schools. However, the most important acknowledgements must go to the 27 teachers who allowed us to learn so much from them.

TABLE OF CONTENTS

Chapter Page

Introduction 1

Organization of the report 3

1 Background and Methodology 5

Development of the instructional model 5

Getting and maintaining the students' attention (Principles 1 & 2) 6

Introducing the lesson and new material to the students (Principles 3,4,5 & 6). 7

Calling on individual students in the group (Principles 7,8,9,10,11 & 12). 10

Dealing with individual learning rates within the group (Principles 13,14,15 & 16) 13

Responding to answers that are not correct (Principles 17,18 &19) 16

Responding to correct answers (Principle 20). 20

Praise and Criticism (Principles 21 & 22). 21

Administration of treatment to teachers 23

Classroom observation 24

Testing of students 31

Data Analyses 33

2 Analysis of Achievement Data for Treatment Effects 35

Analyses using class mean readiness as a covariate 35

Other analyses 39

Summary of analyses of treatment effects of class achievement 42

3 Results of Group Comparisons 45

Getting and Maintaining the students' attention (Principles 1 & 2) 47

Summary (Principles 1 & 2) 50

Chapter Page

3 (cont.) Introducing the lesson and new material to the students (Principles 3,4,5 & 6). 51

Summary (Principles 3,4,5 & 6) 56

Calling on individual students in the group (Principles 7,8,9,10,11 & 12) 57

Summary (Principles 7,8,9,10,11 & 12). 67

Dealing with individual learning rates within the group (Principles 13,14,15 & 16) 69

Summary (Principles 13,14,15 & 16). 70

Responding to answers that are not correct (Principles 17,18 & 19). 71

Summary (Principles 17,18 & 19). 89

Responding to correct answers (Principle 20). . . . 90

Summary (Principle 20). 93

Praise and criticism (Principles 21 & 22). 93

Summary (Principles 21 & 22). 98

Time Usage 99

Summary of time usage data 102

Curriculum used and content covered 103

Summary of group differences on curriculum used and content covered 106

Other categories of academic teacher-student interaction 107

Summary of group differences for other categories of academic interaction 112

Behavior contacts 113

Summary of group differences on behavior contacts 116

Discussion 116

Chapter Page

4 Results of Regression Analyses 119

Data analysis 119

Linear relationships 119

Curvilinear relationships 122

Results 127

Getting and maintaining the students' attention (Principles 1 & 2) 127

Summary (Principles 1 & 2) 130

Introducing the lesson and material to the students (Principles 3,4,5 & 6). 131

Summary (Principles 3,4,5 & 6) 134

Calling on individual students in the group (Principles 7,8,9,10,11 &12) 136

Summary (Principles 7,8,9,10,11 & 12). 145

Responding to individual differences within the group setting (Principles 13,14,15 & 16). 147

Summary (Principles 13,14, 15 & 16) 147

Responding to incorrect answers (Principles 17,18 & 19). 148

Summary (Principles 17,18 & 19). 170

Feedback to correct answers (Principle 20) 173

Summary (Principle 20). 173

Praise and criticism (Principles 21 & 22). 174

Summary (Principles 21 & 22). 177

Time Usage 178

Summary of results of time usage variables . . . 181

Curriculum and content covered 182

Chapter Page

4 (cont.) Summary of regression data for curriculum used and content covered 185

Other categories of academic teacher-student interaction 186

Summary of results of regression analyses of other academic interaction categories. . . . 196

Behavior contacts 196

Summary of regression analyses for behavior contacts 201

5 Summary of Results and Revision of the Instructional Model. . 203

Getting and maintaining the students' attention (Principles 1 & 2) 208

Introducing the lesson and new material to the students (Principles 3,4,5 & 6). 210

Calling on individual students in the group (Principles 7,8,9,10,11 & 12). . . . . 21[illegible]

Dealing with individual learning rates within the group (Principles 13,14,15 & 16) 220

Responding to answers that are not correct (Principles 17,18 & 19). 221

Responding to correct answers (Principle 20). . . . 227

Praise and criticism (Principles 21 & 22). 228

Time Usage 231

Curriculum used and content covered 233

Other response opportunity categories 236

Behavior contacts. 242

Discussion 245

6 Suggestions for Future Experimental Studies of Teaching. . . 251

Maintaining the ecological validity of the classroom in an experimental study 252

Chapter Page

6 (cont.) The content of the treatment 253

Clustering of specific suggestions 255

Maintaining experimental control in a classroom setting 260

Conclusion 267

Reference Notes 269

References 271

Appendix A: An instructional model for first-grade reading groups 275

Appendix B: Summary of the observation system used in the First-grade Reading Group Study 309

Appendix C: Glossary of terms used to describe data from the First-grade Reading Group Study 329

LIST OF FIGURES

Number		Page
1	Example of tables describing the results of regression analyses	124
2	Summary of major results of analyses of group differences (treatment vs. control) and relationships with achievement	206

The First-grade Reading Group Study was an experimental effort developed from the integration of research and knowledge about how young children function in a classroom, especially within the small group format. The most important sources of the ideas in the study were the Texas Teacher Effectiveness Study (Brophy and Evertson, 1976; Note 1), program development work done at the Southwest Educational Development Laboratory (1973) and the work of Marion Blank (1973).

The result of the integration of these sources was an instructional model consisting of 22 specific principles believed to promote effective teaching of young children in small groups. This model is presented as it applies to first-grade reading groups in Brophy, Anderson, Greenhalgh, Odgen, and Selig (Note 2), and in Appendix A.

Although the ideas present in the instructional model are based on previous research and experience suggesting their effectiveness in producing student learning, the purpose of the study was to test the model experimentally to confirm this. Such experimental efforts are necessary if the findings of correlational studies are to be validated and relationships between variables explained.

The past several years have seen many process-product studies of teaching effectiveness, in which observed teacher behaviors are related to student outcomes, most usually achievement. For general reviews of work in this tradition, see Rosenshine (1976), Borich (1977), and Medley (Note 3).

These process-product studies have been correlational in nature, so that it has not been possible to say that a set of teaching behaviors directly led to student outcomes. Rosenshine and Furst (1973) described three important stages in the design of research on teaching: descriptive, correlational, and experimental, and they emphasized that the last step must be taken before relationships between classroom process variables can be accepted as valid, because only experimental results will allow conclusions about causality. Dunkin and Biddle (1974) echoed this by noting, after an exhaustive survey of existing research on the effects of classroom processes, that "process-process and process-product experiments concerning teaching should be encouraged, but preferably for the validating of crucial relationships previously discovered in field surveys or with strong theoretical justification" (p. 466).

It was from this perspective, therefore, that the First-grade Reading Group Study was conceived and carried out. There existed a body of information about teaching practices in a particular setting (small group instruction in early elementary grades). These were integrated into an instructional model consisting of 22 specific principles describing teaching techniques which served as the "treatment" in an experimental study. Extensive observational data were collected in 20 first-grade classrooms to allow the investigators to examine the effects of the treatment on both student achievement and teacher behaviors, and to examine process-product relationships, including some directly relevant to the treatment and others independent of it.

This report presents and discusses data that address the major questions of the study at the classroom level of analysis. Because the report contains much detail and addresses many questions, the reader is advised to review the outline given below and consider the suggestions before beginning to read.

Organization of the report. The report is divided into six chapters. With the exception of Chapters 2 and 6, the text is organized around categories that correspond to parts of the instructional model and other observational variables.

Chapter 1 presents the background and methodology of the study, and discusses the instructional model.

Chapter 2 presents data on treatment effects on student achievement.

Chapter 3 presents data on group differences on process measures in order to determine if the treatment teachers' behaviors actually differed from the control teachers'.

Chapter 4 presents process-outcome data, in order to determine how teaching behaviors were related to student achievement.

Chapter 5 is a discussion and contains suggestions for revising the instructional model for future use, given the results of this study.

Chapter 6 contains suggestions for the design of future experimental studies.

Within Chapters 3, 4 and 5, detailed results are presented and/or discussed according to eleven categories of variables, and summaries of results for these are provided at the end of each section in Chapters 3 and 4. The categories of variables are:

1. Getting and maintaining the students' attention (variables derived from Principles 1 and 2 in the instructional model);
2. Introducing the lesson and new material to the students (Principles 3, 4, 5, and 6);
3. Calling on individual students in the group (Principles 7, 8, 9, 10, 11, and 12);

4. Dealing with individual learning rates within the group (Principles 13, 14, 15, and 16);

5. Responding to answers that are not correct (Principles 17, 18, and 19);

6. Responding to correct answers (Principle 20);

7. Praise and criticism (Principles 21 and 22);

8. Variables describing time usage;

9. Variables describing curriculum and content covered;

10. Variables describing other categories of academic teacher-student interactions (response opportunities) not discussed in the instructional model;

11. Variables describing behavior contacts.

The reader who is most interested in general patterns, but not in results for specific variables, should read Chapter 1 and the discussions (Chapters 5 and 6), and may also want to look at the summary sections in the other chapters. Readers who are only interested in certain variables may examine them by moving to the pertinent sections (as listed above) in Chapters 3, 4, and 5.

Data for all analyses are presented in the tables in Volume II. Most of these tables are reduced copies of the computer printout, and the order of variables in the tables does not reflect the order in which the variables are discussed. Therefore, pertinent statistical information is provided in the text regarding the strength of significant findings, means, and ranges. However, variable numbers are provided so that the reader can look up other information provided in Volume II.

Chapter 1: Background and Methodology

There were four important stages in planning and collecting data for the First-grade Reading Group Study. Each is discussed in this chapter to give the reader the necessary background to understand results presented in other chapters.

1. Development of the instructional model. The model was the basis of the study, and was the result of integration of past research on effective teaching of young children. It was described in a short booklet that was given to teachers in the treatment groups.

2. Selection of teachers and administration of treatment. Twenty-seven teachers were involved in the study, divided into three groups: treatment-observed, treatment-unobserved, and control. Teachers in the two treatment groups were given the booklet describing the instructional model, and they agreed to implement the model's principles in their teaching.

3. Classroom observation. Throughout the school year, teachers in the treatment-observed and control groups were observed teaching their reading groups, and data were collected with a coding system which had been designed to measure implementation of the instructional model.

4. Testing of students. At the end of the school year, the students in the 27 classrooms were given the reading subtests of the Metropolitan Achievement Test, Level I. These students' scores on the Metropolitan Readiness Test (given at the beginning of the year) were used as covariates in analyzing their achievement.

Development of the Instructional Model

The model was presented to the teachers as a set of guidelines for teacher management of reading group instruction. It was "curriculum free" in that it did not focus on the content or materials used in teaching reading, but only on teacher behaviors involved in managing the group as a whole or managing individual student responses. A major rationale for the model was that every child should receive as much individual attention as possible within the group setting, and a major objective of the model was to help teachers achieve the optimal balance between attention to the group and attention to individuals.

It was emphasized that the teacher's role in implementing the model was an active and important one. Application of the principles of the model involved teacher judgment based on knowledge of individual students' needs and the group's needs. The principles were meant to serve as guidelines to be applied as each teacher thought best for each of her groups (as it happened, all 27 teachers were female).

The background and rationale for each principle are discussed below. The first 16 principles have to do with organization and management of the group as a whole, and the rest concern teacher responses to individual student answers. They are grouped as they were presented to the teachers. In the manual given to the teachers, the presentation of each principle (denoted here by underlining) was followed by a brief discussion of the rationale and some practical examples (see Appendix A).

Getting and Maintaining the Students' Attention (Principles 1 & 2)

Principles 1 and 2 emphasized that it is important to catch and maintain the children's attention at the beginning of the lesson.

1. The teacher should use a standard and predictable signal to get the children's attention. In discussing this principle, it was suggested

to the teachers that they use standard attention-getters in two situations: when engaged in transitions from general class activities to the reading group, and when getting students' attention at the beginning of the group lesson. The rationale for this principle was that less time would be wasted in transitions and in "settling down" behavior if the students learned to respond "automatically" to a familiar signal. This technique was based on research suggesting that good management systems minimize wasted time and opportunities for disruptive behavior (Kounin, 1970).

2. Once in the group, the children should be seated with their backs to the rest of the class while the teacher is facing the class. The rationale for this principle was that the students in the group would be less likely to be distracted by other activities in the class when seated this way, and that the teacher would be better able to monitor activities in the rest of the room while teaching the small group. Kounin (1970), discussed the importance of classroom monitoring as an aspect of "withitness" and as a way to prevent problems. The teacher who is working with a small group remains responsible for the rest of the class, and therefore must prevent disruption from outside the group if the reading group lesson is to be taught effectively, and if out-of-group students are to spend their time productively.

Introducing the Lesson and New Material to the Students (Principles 3,4,5, & 6)

Principles 3, 4, 5, and 6 were concerned with introducing new material, and were based on the premise that an introduction should prepare the students for the lesson by getting their attention, teaching new skills and terms before asking the students to apply them, and making sure that the students know what to do in activities.

3. The introduction should contain an overview of what is to come in order to mentally prepare the students for the presentation. The rationale for this principle was that students who are "mentally prepared" for new knowledge or future activities can better receive and process that information. That is, an overview should help students organize their thinking and focus on the task at hand by pointing out relevant aspects. Research support for this principle came from Brophy and Evertson (Note 1), who found a weak but positive relationship between use of advance organizers at the beginnings of lessons and student achievement in higher-SES classrooms. More general support for the principle came from Ausubel (1963), who emphasized the importance of advance organizers.

4. It is also at the beginning of the lesson that new words and sounds should be presented to the children so that they can use them later when they are reading or answering questions. The rationale behind this principle was that students who know what to expect later in the lesson will be able to practice new skills more easily than they would if they encountered unfamiliar words in the midst of reading. It was assumed that reinforcement of new words within the reading lesson would be greater when the words were presented at the beginning, or at least prior to encountering them within some context. Therefore, a rationale for this principle was that breaking up new information into introductory and practice phases would make it easier for the students to learn it.

5. When new words or sounds are presented, the teacher should have the children repeat them until they can say them satisfactorily. This principle was an extension of Number 4; that is, once new information (in this case, new words) has been presented, it is important to initially practice using that information in small, simple steps so that the students gradually increase their skills in using it. (This is presumed to be especially important with "tool skills" such as beginning reading.) Research support for this principle came from Brophy and Evertson (Note 1) who found that practice of new material was positively related to learning in high-SES schools.

6. After moving into the lesson, but before asking the children to use new material or undertake new tasks, the teacher should present a demonstration and/or explanation of any new activity. The discussion of this principle emphasized that a good demonstration or explanation included a carefully sequenced presentation of the processes involved in completing an activity, and was given in simple, clear language that children could understand. However, it was also emphasized that the teacher is the best judge of how much detail and how many steps need be included in an explanation for a given group or student. Research support for this principle came from Brophy and Evertson (Note 1), who found that, in general, lower-SES second- and third-grade students, and therefore presumbly lower ability students or those at the beginning of basic skill learning, benefited from carefully sequenced instruction, with enough redundancy of information to insure that students understand.

Calling on Individual Students in the Group (Principles 7, 8, 9, 10, 11, & 12)

Principles 7 through 12 dealt with calling on children. This involved distributing individual response opportunities during the lesson, while at the same time keeping the entire group alert.

7. The teacher should work with one individual at a time in having the children practice the new skill and apply the new concept, making sure that everyone is checked and receives feedback during the lesson. The rationale behind this principle was that the teacher needed to monitor the progress of each group member, and that the only way to do this was to question each child individually. This implied that excessive use of choral responses would not be desirable. Research support for this principle came from Brophy and Evertson (Note 1) who found that second- and third-grade teachers who spent much time with individual students, even within a group context, were more successful than teachers who tried to work with the class as a whole or with pairs of students. This result was interpreted to indicate that students at these grade levels need individual monitoring and feedback, especially when learning new material and trying to apply it for the first time. A consequence of this principle was that excessive use of choral responses should be avoided, since this might mean that some individuals were not genuinely practicing a skill, and it would be difficult for the teacher to recognize this from choral responses.

8. The teacher should use a pattern (such as going from one end of the group to the other) for selecting children to take their turns reading

in the group or answering questions (as opposed to calling on them randomly and unpredictably). The rationale for this principle was that students would know when to expect their turn responding, and that this would result in both lowered anxiety about being called on unexpectedly and increased teacher control of overeager students who tend to call out answers or volunteer more intrusively than the quieter students. Research support for this principle came from Brophy and Evertson (Note 1), who found that teachers who did not use systematic procedures for selecting students produced lower learning gains. This principle is related to #7, which emphasized the need to work with individuals so that all children could practice skills and receive feedback. By use of systematic selection, the teacher insures regular contact with each student to practice important skills.

9. *In order to keep each member of the group alert and accountable to all times between turns, the teacher should occasionally question a child about a previous response from another child.* The rationale behind this principle was that the occasional use of such comments would prevent any lapses of attention that might arise from the use of ordered turns. It was felt that these two techniques used together would produce optimal attention, as well as the other advantages of ordered turns described above. This technique was suggested by work done by Southwest Educational Development Laboratory (1973), where it was included in a series of staff development materials for working with kindergarteners in small group settings, and in turn was based on the work of

Kounin (1970).

10. Calling on volunteers should be primarily restricted to parts of the lesson in which children are contributing personal experiences or opinions. The rationale for this principle was that teachers who relied too much on volunteers would not be distributing response opportunities equally, so that shyer students might have less contact with the teacher and less skill practice than they needed. It was felt that some situations probably were appropriate for using volunteers, such as giving personal experiences or opinions, so it was left up to the teacher to decide when volunteers should be used. However, the principle emphasized that the best way to achieve the objectives of lessons focusing on reading skills was to use ordered turns and occasionally question a student out of turn. Research support for this principle came from the findings cited above (#7 and #8) for the use of systematic selection and the value of interacting with each individual student as much as possible.

11. When call outs occur, the teacher should remind the child that everyone gets a turn and he must wait his turn to answer. The rationale behind this principle is similar to that of #10, in that letting students call out answers often results in the quieter, shyer students getting less interaction with the teacher. Research support for this principle comes from Brophy and Evertson (Note 1), who found that the frequency of call outs had negative relationships with learning gains in high-SES schools. There were neutral and positive relationships for this behavior in low-

SES schools, suggesting that call outs might be an index of student motivation and enthusiasm. Therefore, in discussing the principle, it was emphasized that the teacher should not be overly critical of call outs, especially when they might indicate student enthusiasm.

12. The teacher should avoid rhetorical questions, asked for effect with no answer expected, or leading questions. Other questioning patterns to be avoided are answering one's own questions, and repeating questions. The rationale behind this principle was that it is important for the teacher to communicate to the students that every teacher question demands an answer, and that questions can be answered through application of skills. It was felt that teachers who used these questioning patterns too much might confuse the students or teacher them to "second-guess" the teacher by responding to her tone of voice or sentence pattern, rather than listening to the content of the question. Support for this principle was based on the authors' observations of teachers and knowledge of how young children may respond in confusing situations, and on the writings of Groisser (1964) and Laughlin (1961).

Dealing with Individual Learning Rates Within the Group (Principles 13,14,15, & 16)

Principles 13 through 16 were concerned with meeting individual learning needs within the group setting. These principles suggested techniques such as breaking up the group, using another child as a model for the group, and arranging for tutorial help for students who were not meeting learning objectives within the standard group setting and time.

13. At some point during the lesson, the teacher must make a fundamental decision about whether the group as a whole can or cannot meet a lesson's objectives. The rationale behind this principle was that

teachers who remained aware of individual differences in rates of learning of new material would be more likely to prevent problems that might arise when one or two students in the group were not learning as desired. In such a case, if the group remained together and the teacher taught at the level of most of the students, these few would be left behind. On the other hand, if she worked with one or two students who needed extra help, the other students would not be spending instructional time efficiently. This principle was suggested by work done at Southwest Educational Development Laboratory (1973). It emphasized that the teacher should make sure that everyone in the group has met each lesson objective before going on to a new step, but that doing this sometimes would mean having to break up the group to focus on a few individuals, to avoid "losing" them.

14. If the teacher decides that the group as a whole cannot reach the objectives at the same time, because of large individual differences in comprehension of the material, she should teach the more able students through to the end of the lesson, dismiss them, and keep in the group those few who need extra help. This principle suggested specific techniques for breaking up the group when the teacher felt that it was wise to do so. It emphasized that the teacher should handle the breaking of the group without fanfare and without negative statements regarding the students who remain for extra help. This principle was also based on the staff development materials developed at the Southwest Educational Development Laboratory (1973).

15. Sometimes the teacher may wish to use one or more children who have mastered the objectives to serve as models for the others. One rationale for this principle was that sometimes students learn more quickly or pay more attention to peers whom they respect and like. Another rationale was that for some types of learning, especially rote or memory skills, it would be less frustrating for the teacher and a student to model the skills rather than to continue working with students who have not learned them yet because they haven't had enough exposure. This principle was suggested by materials produced by the Southwest Laboratory (1973). They recommended this especially for skills used in learning a new language, or other very basic material that needs to be overlearned to become automatic. They suggested that students who were learning new language patterns need the opportunity to hear correct patterns being spoken more than opportunities to practice speaking them, at the beginning. It was felt that this was applicable to teaching beginning reading, because many of the skills involved there also need to be overlearned to the point of automatic response.

16. If one or more children still do not succeed in meeting the objectives within the time available for the lesson, provision should be made for tutorial assistance. The rationale for this principle was that students who failed to meet objectives within a group lesson needed to receive extra help instead of being allowed to fall behind. That is, the reading group setting could not be effective for them once they got behind and had missed important skills that were assumed and built on in succeeding lessons. This principle follows directly from

the fact that the beginning reading curriculum is structured hierarchically. It also is compatible with the principles of programmed instruction and of mastery learning (Bloom, 1976).

Responding to Answers That Are Not Correct (Principles 17,18, & 19)

The second part of the model was concerned with the teacher's role in dealing with individual students in the group. These principles focused primarily on the feedback given to students about their individual answers. Teacher judgment was especially crucial for these principles, because they distinguished among types of questions, types of pacing, and types of student answers:

a) A distinction was drawn between two types of questions: those which called for short factual answers requiring only memory, and those which could be reasoned out. Giving students hints and encouraging them to reason through to an answer is a possible and sometimes desirable tactic to be used in connection with the latter type of question, but not the former. Factual questions usually require factual feedback.

b) It was assumed that different learning objectives would require different pacing strategies. Some lesson objectives are taught most effectively using fast-paced drill and short answers, while others are taught better in slower paced lessons. Extended feedback from the teacher takes time, so that the pace of the lesson is important to consider when offering feedback.

c) Obviously, the quality of the student's answer is very important to consider when deciding on feedback. Information given to a child about a correct response will differ from that given about an incorrect response. The problem facing a teacher when a child does not respond at all is very different from the problem of reacting to a partly incorrect response. Each of these situations requires a different feedback response from the teacher, depending upon the demands of the question and the capability of the child.

In general, the last six principles were based on the premise that any child's response, whether correct or incorrect, could be turned into a pleasant learning experience by the teacher, using appropriate feedback that considered both informational needs (types of question and types of answer) and the lesson's pace.

17. After asking a question, the teacher should wait for the child to respond and also see that other children wait and do not call out answers. During rapid pacing, she should wait a few seconds and give the answer if there is no response. During the more slowly paced parts of the lesson, the teacher should wait for an answer as long as she feels the child is thinking and will answer, but not so long as to embarrass the child or lose the other children's attention. If the child does not respond within a reasonable time in slower paced lessons, the teacher should indicate that some response is expected by probing. She should then simplify according to Principle 19.
    The rationale for this principle was that students should learn that a response is expected of them, and that the teacher should encourage this whenever she can without disrupting the pace of the lesson. Research support for this principle came from Brophy and Evertson (Note 1) who found that teachers, especially low-SES teachers, who had lower rates of "no responses" had higher learning gains. The work of Marion Blank (1973) also suggested this principle. Her guidelines for teaching are derived from a program for disadvantaged pre-

school children and centered on dialogues between a teacher and a child, in which the teacher's handling of an incorrect response or a lack of response is considered crucial. Generally, the suggestions made in this instructional model about feedback to incorrect answers correspond to Blank's suggestions.

18. When the child is incorrect, the teacher should indicate that the answer is wrong, and then follow simplification procedures outlined in Principle 19. In communicating this principle to the teachers, it was emphasized that the incorrect answers should not be met with overly negative or rejecting reactions by the teacher, but that the student should know clearly what was wrong about the answer. The teacher should try to be as specific as possible about what was wrong. The rationale for this principle was that the student needed informative feedback if the incorrect answer was to be used constructively. Research support for this principle is discussed below under Principle 19.

19. The appropriate simplification procedure is determined by the type of question.

    a) If the question deals with factual knowledge that cannot be reasoned out, the teacher should give the answer to the child and then move on.
    b) If the question is one that the child could reason out with help, the teacher should provide clues or simplify the question. If clues still do not help the child, he should be given the answer. The teacher should never ask another child to supply the answer.

In explaining this principle to the teachers, it was emphasized that it was much more important for the teacher to stay with the child who has answered incorrectly, or who has failed to respond, than to go on to another student to get the answer. When the teacher gave the answer to the child, it could be done in several ways. The teachers were told that this depended on the pace that they were trying to maintain in the lesson. If it were rapid, the teacher should probably give the answer and move on, perhaps occasionally having the child repeat the response. If the pace were slower, the question could be restated in a form that simply called for agreement, repetition, or choosing between alternatives. When the question was one that the student was expected to figure out (with help if necessary), the teacher could give clues or rephrase in ways that guided the child's thinking in the right direction. If these clues did not help, the teacher could then give the answer (rather than call on another child).

The rationale underlying this principle was that first graders, at this point in their learning of basic reading skills, were more likely to listen to and understand information given during direct interactions with the teacher than they were to learn when hearing another student give the answer. Also, there might be unfortunate affective consequences if teachers regularly gave up and moved on to other students when the first student didn't answer. A teacher who relied instead on the "sustaining" approach embodied in this principle feedback was likely to communicate to each student that she expected and would be able to elicit some acceptable response to each question, and that all students could learn to listen, think and respond.

28

Research support for this principle came from the findings of Brophy and Evertson (Note 1) for low-SES students in second and third grades. For these students, there were positive relationships with learning for teacher use of sustaining feedback, but negative relationships for asking other students for answers. This behavior was not as important for high-SES students. This was interpreted to mean that students who are still learning basic tool skills (as first graders would be, and as the lower-SES second and third graders were) benefit more from sustained interaction with the teacher than from listening to their peers. Another source of support was Blank (1973) who described several simplification techniques useful in certain situations, depending on the characteristics of the question, the child's response, and the child's personality. A program based on her thinking was successful in improving performance on IQ tests of disadvantaged pre-schoolers (Blank and Solomon, 1968). Discussion of this principle in the teacher materials also included suggestions for giving explanations when supplying answers to students, rather than giving just the answers themselves.

Responding to Correct Answers (Principle 20)

20. When the student has answered correctly, the teacher should acknowledge the correctness and make sure that everyone else heard and understood the answer. The rationale for this principle was that young students do not necessarily know when they are correct, and that they deserve informative feedback on this point. It is also important for other students in the group that the correct answer be acknowledged. It was

suggested in this principle that if other students did not hear or understand the answer, then the teacher might repeat the answer or have the original student repeat it. However, it was suggested that the teacher not get in the habit of following every answer with repetition. There was no direct, specific research to support this principle, but its rationale stemmed from recognition of the source of young students' misunderstandings of what is and is not correct and from general stress on the importance of feedback.

Praise and Criticism (Principles 21 and 22)

21. Praise should be used in moderation. The teacher should praise thinking and effort more than just getting the answer, and should make praise as specific and individual as possible. The rationale here was that praise should be used on an occasional basis to reinforce the students, but if used too much, it would lose its value. It was assumed that making praise as specific as possible would convey more information to the student about his answer, and would therefore be more effective feedback. The research base for this principle was a set of sometimes contradictory findings. Although there is much agreement that praise can be an effective motivator, there has been little evidence that there is a strong relationship between learning or good performance and verbal praise. It was hypothesized that one explanation for this lack of relationship was that most praise was not given very effectively. Therefore, it was the purpose of this principle to try to make praise more effective and meaningful, and to examine its effect, given these characteristics.

22. Criticism should also be as specific as possible, and should include specification of desirable or correct alternatives. The rationale for this principle was that there is sometimes a reason to give criticism because it can be informative to students and can point out the relative aspects of their behavior and/or thinking. However, it was felt that the more specific the criticism, the more information is presented to students. Therefore the rationale for this principle was very similar to that for praise. The research base for this principle came primarily from Brophy and Evertson (Note 1), who found that the more effective high-SES teachers used some criticism. When used by these teachers, it was a reflection of high expectations and demandingness rather than negativistic responses to the student's work. In the principle as it was explained to the teachers, it was emphasized that specific, positive instruction should be given along with any correction or criticism.

In summary, these 22 principles created an instructional model which had as its underlying rationale an emphasis on getting and maintaining students' attention, sequencing information clearly for the students, and being very careful to provide information about the relevant aspects of a question or answer. Although it was not expressed this way in the materials given to teachers, the model clearly suggests that the teacher play a controlling and leading role in directing the reading group. In this sense, the instructional model can be said to be a reflection of "direct instruction" (Rosenshine, 1976), in that it asks the teacher to take on the role of instructional leadership, through constant monitoring and control of students' behavior and information processing.

Administration of Treatment to Teachers

After developing the instructional model, the next step in the study was to give it to teachers in the treatment groups and arrange for a control group who did not receive treatment. Nine elementary schools and 27 female first-grade teachers were involved in the study, divided among three groups as follows:

1. Treatment-observed. Ten teachers in three schools received the treatment (were instructed in the principles and agreed to use them in their teaching) and were observed teaching each of their reading groups throughout the year. (Student N = 192)

2. Treatment-unobserved. Seven teachers in three schools received the treatment but were not observed during the year. This group was included to assess treatment effects in the absence of observation. (Student N = 147)

3. Control-observed. Ten teachers in three schools were given no special instructions about how to teach. They were observed throughout the year. This group was included in order to measure natural implementation of the principles in the absence of a treatment. (Student N = 218)

The schools were assigned to treatment groups by first creating three groups of three schools each which were comparable in SES composition. Although all of the schools were located in neighborhoods which were predominantly middle class and Anglo, there was some slight variation among the schools in SES ratings assigned by the district. Therefore, three groups were created so as to be balanced in this respect, before being randomly assigned as treatment or control groups.

The result of this process was that all participating teachers within a school were assigned to the same treatment group. This opened the possibility of a school effect in the results, but this was considered a less serious risk than the possible contamination that would occur if teachers within a school were assigned to different groups. In evaluating the overall effect of the treatment on achievement, the possibility of a school effect was considered. These analyses are discussed later in the report.

All teachers who participated in the study had agreed to do so after discussing it with the principal investigators. Teachers in the two treatment groups were told the purpose of the study (i.e., to experimentally test earlier correlational findings). Teachers in the control group were told that the purpose of the study was to find out more about effective teaching of first-grade reading.

The 17 teachers in the treatment groups were given a short booklet (33 pages) which described the instructional model by presenting each principle and its rationale (see Appendix A). They were asked to read it and meet again a week later with one of the investigators, to discuss any questions they had. At this second meeting, the teachers took a short, multiple-choice test over their knowledge of the principles. All treatment teachers demonstrated sufficient knowledge of the model, and this was the extent of the treatment, although the teachers kept the booklets for reference. They each agreed to implement the principles in the model as they deemed them appropriate for their reading groups.

Classroom Observation

Data Collection

The treatment was applied in October, 1974, and observations of teachers in the treatment-observed and control groups began in November. From this time until May, 1975, each of these 20 classrooms was visited 15 to 20 times (approximately once a week), and observed systematically with a coding system developed specifically for the study (Brophy, Mahaffey, Greenhalgh, Ogden, and Selig, Note 4). Three of the teachers (two treatment-observed and one control) left their schools after 2 months of observation for reasons unrelated to the study. Their replacements agreed to participate in the study and were observed the rest of the year. Therefore, scores for these teachers are based on about 12 observations.

The observation system was designed with the 22 principles in mind. Therefore, it incorporated measures of implementation of the model, as well as other measures to assess the possible effects of such implementation. The coding system was organized so that it would follow the natural flow of activities during the reading group, but it could be broken down later into specific variables most relevant to discussion of each principle.

The system was divided into two parts. The first focused on the teacher's dealing with the groups as a whole, and the second involved her academic interactions with individuals. This division reflected the theme running throughout the instructional model: the importance of maintaining a balance between management of the group as a whole and interaction with individuals within the group. Appendix B contains a summary of the observation system, and Appendix C provides a glossary of terms derived from it.

Group data collected during the observations. The measures in this section described the teacher's interactions with the group as a whole. These included activities occurring before the group lesson began, as well as certain contacts with the group as a whole that took place during lessons.

The first thing the observer would note during each observation was information about the teacher's managing the transition to the group and getting the attention of the students once they were in the group. Specific measures here were addressed to the types of attention-getters used and the length of time it took to get students to the group and to get the lesson started. (This measured implementation of Principle 1.) At this point, the observer would note how the students and the teacher were seated with respect to the rest of the class (Principle 2). Once the lesson was begun, the observer noted the use of an overview and its effects (Principle 3). At this point, the lesson proper would begin, and the observer would record information about interactions between the teacher and individual students (described below). However, during the rest of the lesson, the observer would note certain information about the way the teacher dealt with the group as a whole, whenever it was available. This included information about breaking up the group (as described in Principles 13 and 14), the use of a student model (Principle 15), the quality of demonstrations and explanations (Principle 6), the presentation of new words (Principles 4 and 5), choral responses and group call outs (Principle 7), and the use of undesirable types of questions directed to the group as a whole (Principle 12).

Individual data collected during the observations. When the teacher started the lesson, the observer began to describe each interaction between the teacher and an individual student that started with an academic question asked by the teacher (response opportunities). Any behavior contacts occurring during the lesson were also recorded. This coding of individual interactions continued until the group was dismissed, although it could be interspersed with coding of information about the group, as described above.

Each response opportunity was described as to the type of selection, the type of question, the type of answer, and the type of feedback. Each of these larger categories included several specific types of student or teacher behavior. For example, under the general heading of "selection," the observer would describe each interaction as to the method of selection used and whether is was an ordered turn, a volunteer, a call out, etc.

Behavior contacts were described in terms of the type of student misbehavior (e.g., social talk, misuse of materials), the type of teacher correction, and whether or not the correction was specific as to desired alternatives.

The data on response opportunities and behavior contacts were used to measure implementation of Principles 7-11 and 17-22.

All of the coding of individual students' interactions with the teacher was "low-inference" in that the observer was classifying specific behaviors into categories according to preestablished definitions. Inference on the observer's part was limited, so that he or she was essentially just counting specific behaviors when they occurred. Some parts of the group data collection were also low-inference, in that they involved counting or timing, but other measures were "high-inference". For these, the observer was asked to rate extent of use of a behavior or degree of appropriateness of its use.

The observers spent two weeks in the classrooms practicing with the system before actually beginning data collection. Pairs observed together until the criterion of 80 percent agreement on each major section of the coding system was reached. After that time, observers worked alone. Each teacher was seen by two observers who alternated visits to her classroom.

Data Preparation

At the end of the year, when all observations were completed, many scores were computed for each teacher to represent the implementation and effect of each principle in the instructional model in her classroom. These scores represented sums for the entire year, standardized where necessary by the amount of time spent in observation. There are three types of scores reported: rates, proportions, and averages.

Rates were computed by dividing a total frequency in a category by either the total amount of time observed (yielding rate per minute) or the number of observations (yielding the average number of occurrences per observation). For the latter computations, an observation was considered to be one complete lesson taught to a single group. Therefore, a single teacher might be observed teaching from one to four groups in a morning, and the data for each observation were considered separately for certain scores.

Proportions were relative rates which were computed by comparing the number of times that an event or behavior occurred to the number of times that it could have occurred. Therefore, proportion scores range from .00 to 1.00. Occasionally, it is easier to discuss a proportion in terms of percentage of time, in which case the corresponding scores are 0% to 100%. Examples of proportions are "the proportion of observations in which the teacher gave an overview of specific instructional content at the beginning of the lesson" and the "proportion of all response opportunities which included correct answers from the child."

Averages are used to discuss ratings or counts taken on a regular basis. These were based on the number of times the scores were reported for a particular variable, which ranged from a few observations during the year to several instances in one observation. Examples of such averages are "the average number of times per observation that the teacher called for a choral response" and the "average rating (on a five point scale) of the percent of students attending to a signal for transition."

In creating variables from the observation system to be used to evaluate the instructional model, several approaches were taken. In many cases, single frequencies were expressed both in terms of absolute rates per minute of time (or per observation) and also as a proportion of the maximum possible score. The resulting list of variables is lengthy but thorough.

The variables created from the low-inference data seem complex at first sight, due to attempts to make each one as fine-grained as possible so as to detect any contextual effects of types of questions, types of answers, and oral reading vs. questions asked outside of reading turns. The sequence used to create these low-inference proportion variables follows.

First, the proportion of all recorded responses ("R.O.'s") that included each separate element was computed, creating variables such as "proportion of R.O.'s that were word recognition questions" and "proportion of R.O.'s with correct answers." Then, the data were tallied separately for finer-grained analyses according to types of questions and/or types of answers, resulting in such variables as "proportion of R.O.'s which were word recognition questions that led to correct answers." (In the tables, this is abbreviated for labeling purposes as "proportion of word recognition questions that were correct." The tables are reduced copies of computer printout, so that the labels had to fit an 80-character limit.) For some variables, a third element was also included, as in, for example, "proportion of reading questions with incorrect answers which received criticism feedback."

In addition to breaking down the variables according to the major division of the low-inference part of the observation system (selection, question, answer, and feedback), all variables describing response opportunities were computed for total interactions, interactions occurring during reading turns (oral reading), and interactions occurring outside of reading turns. This was done because oral reading and other questions were coded differently due to the different contexts involved. When a child was reading aloud, the observer coded information only when the child had an interaction with the teacher during the turn. This was almost always due to an error in reading. If the child read completely correctly, the observer would note separately that a turn had been completed, but would code no other information, because no interactions with the teacher would have occurred. However, when a child was asked a single question by the teacher (not in a reading turn), the observer always completed a line of coding which described the type of selection, question, answer, and feedback.

Therefore, the coding data for each teacher contained information about the ways she dealt with interactions during reading turns (which were almost always errors) and the ways she interacted with students outside of reading turns when asking them single questions (where the error rate varied). There were many commonalities across the two situations, but there were also some important contextual differences. Therefore, separate low-inference variables were computed for each of these two different kinds of interaction, and these two also were combined into data for "total" interactions. This led to series of variables such as "proportion of total R.O.'s which contained praise," "proportion of turn response opportunities which contained praise," and "proportion of nonturn response opportunities which contained praise." It was hoped that such divisions would yield information about what factors must be considered in each setting when making decisions about selection, level of difficulty of material, types of feedback, and the effects of such behaviors on overall pace.

Each score was computed for each teacher, representing her average behavior across all students and all observations.

Testing of Students

At the end of the school year, the observers administered the Metropolitan Achievement Test, Level I, to the students in the 27 classrooms. Only the reading subtests were given. (Until this time, the observers had not interacted with the students.) The teacher assisted in test administration when necessary, so that a ratio of one adult to about 15 students taking the test was maintained in all classrooms. However, the directions were always given by the observer, according to the test-makers' instructions (Durost, Bixler, Wrightstone, Prescott, and Balow, 1970). The tests then were scored according to the manual and standard scores for each student were computed from raw scores according to tables provided by the test manufacturer.

The reading subtests were:

1. Word Knowledge--The students looked at a picture of a familiar object and selected one of four printed words which named it.

2. Word Analysis--The students listened to the test administrator read a word, a sentence containing that word, and the word again, and then selected the word from a group of four printed for that item number.

The words in each group of four were always similar in appearance.

3. Reading--The students read sentences or paragraphs, and then read questions about the material, selecting the appropriate answers from three or four choices.

In addition to these three scores, a Total Reading score was computed by summing the raw scores for Word Knowledge and Reading. (This was done according to Metropolitan scoring instructions.)

The students who had attended school at the beginning of the year had been given the Metropolitan Readiness Test by their teachers. These scores were made available to the researchers, and the Total Readiness score was used as a covariate in all analyses involving student achievement. The Total Readiness score was computed from six subtests:

1. Word Meaning--The student selected one of three pictures which illustrated a word given aloud by the teacher.

2. Listening--The student selected one of three pictures which illustrated phrases or sentences read aloud by the teacher.

3. Matching--The student marked one of three pictures which matched most closely a given picture.

4. Alphabet--The student marked one of four lower-case letters which was read by the teacher.

5. Numbers--The students followed directions read by the teachers which tested number knowledge, such as marking the box with seven dots.

6. Copying--The student was to copy designs, letters, or numbers.

The manual accompanying the Readiness test gives these correlations (N = 743) between the Total Readiness score and the three reading subtests of the Achievement Test, Level I, when given in April of the first grade (Hildreth, Griffiths, and McGauran, 1969):

Word Knowledge: $r = .68$
Word Analysis: $r = .66$
Reading: $r = .64$

In the present sample, class means were computed to use in analyses of other data at the class level. These correlations were found between the Total Readiness score and each achievement subtest ($N = 20$).

Word Knowledge: $r = .75$
Word Analysis: $r = .71$
Reading: $r = .58$
Total Reading: $r = .66$

These correlations approximate those expected on the basis of the test manual, with none of the differences even approaching statistical significance. Class means were computed for each readiness and achievement subtest. Table 1 (in Volume II) provides means of readiness and achievement test scores for each of the three groups of classes.

Data Analyses

A variety of analysis procedures were used to examine the data in the study. These are summarized below and described in more detail in Chapters 2, 3, and 4. All analyses were performed using class means, so that the class and/or teacher is the unit of analyses, yielding an N of 20 for analyses involving observation data and an N of 27 for analyses involving only achievement data. (Seven treatment teachers were not observed.)

Three important questions were addressed with the data: 1) Did the treatment have an effect on student achievement? (Regression analyses of student achievement scores were performed using student readiness and treatment group membership as predictors. These data are discussed in Chapter 2.); 2) Did the treatment have an effect on behaviors? (Analyses of variance were performed on the observation data using treatment group as a classifying variable. These data are discussed in Chapter 3.); 3) What were the relationships between behaviors and student achievement? (Regression analyses of student achievement were performed using variables from the observation system and student readiness as predictors. These data are discussed in Chapter 4.).

Chapter 2: Analysis of Achievement Data for Treatment Effects

This chapter examines treatment effects on achievement to determine if class mean reading achievement was affected by the teacher's membership in one of the three groups.

Several analyses were performed using different covariates. The most important of these analyses involved looking at class mean achievement as a function of class mean readiness and treatment group membership in order to determine if there were differences in achievement between the treatment classes and the control classes when entering readiness was taken into account.

Analyses Using Class Mean Readiness as a Covariate

These analyses were done by a series of linear regression models, according to procedures discussed in Ward and Jennings (1972). This approach to regression analysis compares pairs of linear models which describe different relationships among the variables. The first of each pair is considered the "full" model, and always accounts for more of the variance in the criterion because it contains more predictor terms. The second model is considered the "restricted" model. The formation of each model is determined by the hypotheses that are being tested. In order to determine how important is the difference between two models, their respective error sums of squares are computed and compared. These are the squared summed differences between the predicted and obtained criterion scores, using the regression coefficients obtained for each model to calculate the predicted scores. The error sums of squares of the two models are then compared by means of an F-test. Significant differences indicate that there is a difference in predictive power between the two models, and, therefore, that the full model is a better representation of the data than the restricted model.

The following models were created to describe relationships among predictors for each of the four achievement test scores:

1. $Y= b_1G_1+ b_2G_2+ b_3G_3+ b_4(R*G_1) + b_5(R*G_2) + b_6(R*G_3) + E_1$
2. $Y= b_7R + b_8G_1 + b_9G_2 + b_{10}G_3 + E_2$
3. $Y= b_{11}R + E_3$
4. $Y= b_{12}R + b_{13}(G_1+G_2) + b_{14}(G_3) + E_4$
5. $Y= b_{15}R + b_{16}G_2 + b_{17}(G_1+G_3) + E_5$
6. $Y= b_{18}R + b_{18}G_1 + b_{19}(G_2+G_3) + E_6$

where:

Y= class mean score on one of the four achievement tests.

$b_{\_}$= regression coefficients for corresponding predictor vectors

R = class mean for the Total Readiness score

G_1= membership in group 1 (scored 1 if class was in the control group, and 0 otherwise)

G_2= membership in group 2, (scored 1 if class was in the treatment-observed group, and 0 otherwise)

G_3= membership in group 3 (scored 1 if class was in the treatment-observed group, and 0 otherwise)

$(R*G_1)$, $(R*G_2)$, $(R*G_3)$= vectors describing the product of the class mean readiness score and group membership for each of the groups. $(R*G_1)$ would be equal to the readiness score if the class was in group 1, and would be 0 otherwise. These vectors were used in testing for interaction between readiness and group membership.

(G_1+G_2), (G_1+G_3), and (G_2+G_3)= vectors containing collapsed scores for group membership. For (G_1+G_2), a class would receive a score of 1 if in either group 1 or 2, and a 0 otherwise. These vectors were used in testing paired comparisons.

E = error terms (composed of the differences between the actual criterion score and that predicted by the equation)

The following comparisons of models were made for each of the four achievement tests:

1. Model 1 (full) vs. Model 2 (restricted) *(Test for interaction)*. This comparison tested for an interaction between entering readiness and treatment group membership. A significant F-ratio indicated that the data are better described with the interactive model (Model 1). No significance indicated that the restricted model without the interaction terms predicted the criterion as well as the more complex full model.

2. Model 2 (full) vs. Model 3 (restricted) *(Test for treatment effect)*. In the event that Model 2 was shown to be as sufficient as Model 1 (by the first test described), Model 2 was then taken as a full model and compared to Model 3 which did not include information about treatment group membership. If this test was significant, it indicated that treatment group membership was an important predictor of achievement in addition to entering readiness. If not significant, then the results indicated that the treatment was not a good predictor of achievement. This is analogous to an analysis of covariance.

3. Model 2 (full) vs. Model 4 (restricted); Model 2 (full) vs. Model 5 (restricted; Model 2 (full) vs. Model 6 (restricted) *(Tests for group comparisons)*. These comparisons were made when Model 2 had been shown to be a better predictor of achievement than Model 3, which meant that there was a "treatment effect." These tests are similar to paired comparisons which would follow an analysis of variance. They were done by creating restricted models which did not differentiate between the two groups (Models 4, 5, and 6) and comparing them to Model 2, which did allow different predictive

weights to be used for each separate group. Significance for any of these tests indicated that the restricted model, which combined two groups, was as good a description of the data as the full model, which did not combine those two groups. That is, achievement of the classes in the two groups was significantly different, when entering readiness was taken into account. Nonsignificance indicated that there were no differences in achievement.

Each of these three types of tests was conducted for each of the four achievement tests, using class mean readiness as a covariate. Statistics for model comparisons are presented in Tables 2-6, including R^2 values, F-ratios, and p-levels, and results are discussed below.

Results for tests of interactions between readiness and treatment group membership. There were no significant results indicating an interaction between entering readiness and treatment group membership for any of the four tests (see Table 2).

Results for tests for effect of treatment group membership. These tests revealed significant differences for each of the four tests. There were weaker relationships for Word Knowledge and Word Analysis ($p = .06$ for each) and more highly significant results for Reading ($p = .02$) and Total Reading ($p = .05$). These results indicate a noninteractive treatment effect on class mean achievement when entering readiness is taken into account (see Table 3).

Results for tests for group comparisons. In order to determine the direction of the group effect, the comparisons were examined. Means and standard deviations for each group are presented in Table 1.

1. Group 1 vs. Group 2 (Control vs. Treatment-observed). These tests revealed weakly significant differences between these two groups for

the Word Knowledge and Word Analysis subtests (p = .07 for each), and more highly significant differences for Reading (p = .01) and Total Reading (p = .05). The means for these two groups showed that the treatment-observed group had higher achievement scores than the control group. According to the test makers' conversion tables, the difference was roughly equivalent to one to two month's grade equivalency (Durost, et al., 1971) (see Table 4).

2. Group 1 vs. Group 3 (Control vs. Treatment-unobserved). There were significant differences found between these two groups for all four subtests: Word Knowledge (p = .03), Word Analysis (p = .03), Reading (p = .02), and Total Reading (p = .02). Examination of the means showed that the treatment-unobserved group had higher achievement scores than the control group (see Table 5).

3. Group 2 vs. Group 3 (Treatment-observed vs. Treatment-unobserved). There were no significant differences between these two groups on any of the four subtests. Examination of the means indicated that the treatment-unobserved group had slightly higher means on all tests, however. (see Table 6).

These data suggest that treatment group membership significantly predicted achievement in addition to information about entering readiness. The two treatment groups had higher achievement than the control group, and there were no significant differences between the two treatment groups. The results support the hypothesis that treatment classes would have higher achievement.

Other Analyses

In order to see if the results could be attributed to the confounding of treatment with schools, other tests were done utilizing information about the schools.

Models were created of the same general form as the ones given on page 36, but which differed in that they described achievement as a function of a) treatment group membership and school mean readiness; b) group membership and both class and school mean readiness; or c) group membership and class size. These analyses were performed on Total Reading achievement scores only.

Results of analyses using school mean readiness. Regression models were the same as those given for the tests involving class means, except that all classes within a school received the same school mean score for the readiness vector (R). When the school mean readiness score was used as a predictor instead of class mean readiness, the pattern in the results remained the same, although the $\underline{p}$-levels were not as low (i.e., no significant interaction or differences between the two treatment groups, but weak or moderate relationships for an overall treatment effect and differences between the control and treatment groups). Results are given in Table 7.

Results of analyses using both school mean and class mean readiness. Class mean readiness correlated higher with class mean achievement than school mean readiness (.69 vs. .39), and it was decided that class mean was a more reasonable covariate. Therefore, in order to test for school effect with as much information as possible about entering readiness, models were created which contained both school and class mean readiness as separate covariates. The results of this series of comparisons were similar to those using class mean alone, except that the $\underline{p}$-levels were lower, and there was a significant interaction between treatment group membership and class mean readiness for the Reading and Total Reading scores ($p = .03$). These data are presented in Table 8.

In order to examine this unexpected interaction, the predicted scores were plotted for each treatment group using the model that contained both class and school mean readiness. The third group (treatment-unobserved) had a relatively restricted range of predicted scores when compared to the other two groups, probably due in part to a lower *N*. When the school mean was added into the equation, it had the effect of truncating the predicted scores even more than with the class mean scores alone, resulting in a near-zero slope for the treatment-unobserved group. This led to significantly different slopes for the three groups. When only class mean readiness was used to predict achievement, the range of predicted scores for the third group was not so truncated, although it was smaller than those for the other two groups. The restricted range occurred in the middle of the ranges found for the other two groups, so that classes in the third group did not represent extreme scores. This, plus the fact that class mean readiness was more highly correlated with class mean achievement, led to the decision to use class mean readiness alone as an ability covariate in other analyses.

Results of analyses using class size. The last series of tests used to analyze the treatment effect compared models which included class size in addition to class readiness. If there were any school effects which were not attributable to differences in entering readiness, they might be due to size of classes, which did vary somewhat from school to school. The Total Reading achievement test scores were analyzed with class size, class mean readiness, and treatment group membership as predictors. With class size included in the models, there was no significant interaction between group membership and readiness, and the treatment effect remained significant ($p = .02$). That is, class size was not contributing substantially to the prediction of achievement, so that controlling for class size did not remove the treatment effect on achievement. Results are presented in Table 9.

Summary of analyses of treatment effects on class achievement. The classes in the two treatment groups had significantly higher adjusted mean reading achievement scores than the classes in the control group, when class mean readiness was used as a covariate. There were no differences in the adjusted mean achievement scores of the two treatment groups, indicating that observation did not moderate the treatment effect. No interactions between entering readiness and the treatment were found.

Several analyses were performed to determine if the confounding of school with treatment was responsible for the treatment effect. School mean readiness and class size were each used as a predictor of achievement. However, they did not contribute to the prediction of achievement, so that a similar pattern of results was obtained as when using class mean readiness and treatment group as the predictors of achievement. Therefore, school effects, at least as measured in these two ways did not account for the treatment effect, except insofar as they may be related to the average class readiness level.

These results indicate that the treatment had a beneficial effect. However, in order to completely evaluate the treatment, the components of it were examined to determine if the treatment teachers were indeed behaving differently from the control teachers in the ways expected, and if those behaviors (as defined by the principles) actually related to adjusted student achievement. Other teacher behaviors not directly related to the treatment were also compared to see if the treatment-observed group differed from the control group in other ways. These questions led to analyses of the extensive set of observation variables that are reported in the next two chapters. Chapter 3 presents comparisons of the treatment and control groups on all observation measures, including implementation of the treatment, while Chapter 4 reports relationships with achievement for all of the observation variables.

Chapter 3: Results of Group Comparisons

Data on implementation of the instructional model in the treatment and control classrooms are presented in this chapter. To determine whether the treatment had any effect on teachers' behaviors, the mean scores for the treatment group on each process measure were compared to the mean scores of the control group in a series of one-way analyses of variance. It was assumed that the scores for the control teachers represented the base rates for these behaviors in the population of first-grade teachers in the area, and that any significant group differences in the expected directions could be attributed to the treatment. Variables measuring direct implementation, indirect effects of implementation, and other processes unrelated to the treatment were compared for the two groups.

The data are presented for groups of related principles in the same order that they were discussed in Chapter 1. For each principle, the hypotheses are presented, the measures used to test them are described, and the results are given in terms of the means of the two groups and the level of significance of the mean differences.

Each variable discussed can be examined more closely be referring to the tables. The numbers in parentheses indicate the variable numbers. As discussed in Chapter 1, the tables were created directly from computer printouts, so that the order of variables in the tables reflects the order within the computer program, rather than the order used to discuss them in the text.

Tables 10 through 12 contain statistics describing these results. The tables are divided according to the type of measure, which may be identified by the variable number as follows: a) Table 10 contains variables numbered 4026-5379, which describe the ways that the teacher dealt with the group as a whole (e.g., calling for transitions, asking for choral responses); b) Table 11 contains variables numbered 1-431, which describe rates per minute of components of interactions between the teacher and an individual student (e.g., rate of ordered selection, rate of correct answer); c) Table 12 contains variables numbered 601-1079, which are proportions describing the relative frequency of types of interactions.

In addition to providing the variable number, so that readers may refer to the tables for more detail, the text also includes some statistical information. When the p-level of a comparison is equal to or below .10, this is reported. The means for each group are also given for many of the variables to indicate the extent of use of each behavior described. When means are reported, they are usually given in parentheses following the discussion of the variable, and always in the same order: *the control mean first, then the treatment mean*.

Three levels of significance were considered in interpreting the results. When $.05 < p \leq .10$, the results are reported, but they should be recognized as weak relationships. When $.01 < p \leq .05$, results are interpreted as representing strong relationships. Because of the relatively low N for these analyses (10 for each group), the probability of Type II errors is high. Therefore it is important to consider results which fell in the range between .05 and .10 to see if they fit into a pattern established by other, more clearly significant variables. No single variable is decisive in interpreting data of this sort. Instead, *patterns* of results are more important.

Getting and Maintaining the Students' Attention (Principles 1 & 2)

Use of an attention getter. Teachers in the treatment group were expected to use attention getters more often to begin transitions and to begin lessons and, as a result, to have quicker and easier transitions, and to get the students' attention at the beginning of the lesson more easily. The use of an attention getter during transitions and at the beginning of the lesson was measured by the proportion of all observations in which some attention getter was used and by information about what kind was used. There were no treatment effects for the proportion of time that attention getters were used in either transitions or at the beginning of lessons (5312). Both control and treatment teachers used attention getters to signal transitions very often. Mean proportion scores were .80 and .88, respectively. However, few teachers in either group used an attention getter to start the lesson (5321; means were .05 and .07).

Although the instructional model did not suggest that any one kind of attention getter was best, there were differences between the groups in the types of signals used in transitions. The control group was more likely (p = .06) to use a bell than was the treatment group (5313; .29 and .05). On the other hand, treatment teachers were more likely (p = .07) to use a verbal signal (5315; .64 and .89). When an attention getter to the entire group was not used, treatment teachers were more likely than control teachers to contact individuals directly (5318; p = .07; .15 and .49). There were no significant differences between the groups for any other types of attention getters (5316, 5319).

Even though no treatment effect was found for the use of attention-getting signals, results for variables describing the effectiveness and smoothness of transitions suggested a treatment effect. For example, treatment teachers spent a smaller proportion of their total observed time in transition to the reading group (4062; p = .02; means = .13 for control and .09 for treatment).

Transitions were divided into three components, and each part was timed. The first component was the time between the first occurrence of some signal and the arrival of the last student in the reading group area. This was known as "transition time to group--students." The second component was the time from the arrival of the last student in the group to the arrival of the teacher, known as "transition time to group--teacher." The last component of the transition was the time from the teachers' arrival to the beginning of the lesson, known as "time to lesson, once group is together." There were no differences between the two groups in the average amount of time spent in each of these components (4026, 4027), but there was a difference in the proportion of the total transition time spent in them. The treatment teachers had a larger proportion of their transition time devoted to getting the students to the group (4059; p = .10; means = .34 and .43). This means that the treatment teachers, on the average, spent almost half of their transition time in student movement, while half was spent in teacher movement to the group and in organizing in preparation for the lesson. The control teachers, on the other hand, spent about 60% of their total transition time in teacher movement and organization. One possible interpretation of these results is that the treatment teachers were devoting less of their allotted time to transitions, and were also better managers of their own movements during the transition, so that proportionately less transition time was taken for teacher preparation. This is possibly reflective of better overall preparation, since fewer problems seemed to be delaying them.

In order to judge the effectiveness of the teacher's signals to the students, the observers were asked to rate on a five-point scale the percent of students who attended immediately to the transition signal (5303) or to a signal given at the beginning of the lesson (5304). There were no differences between the two treatment groups on these measures (means = 3.34 and 3.38 for transitions, 3.74 and 3.72 for lessons).

Another variable measuring the effectiveness of attention getters was the frequency of teacher corrections to students about behavior during transitions or about inattention before beginning the lesson. There was a difference (p = .06) in the average number of corrections given to individuals during transitions (4042). The mean for the control group was 1.29, and for the treatment group it was .80. This probably represents fewer problems in transitions for the treatment group, although it could indicate a reluctance on the part of the treatment teachers to correct students.

The number of times the teacher had to repeat an attention getter was also noted. There were no significant differences in the number of times that signals had to be repeated. Both groups averaged one signal per transition, and neither group used a signal very often to start the lesson (4041, 4043).

Seating the group. Principle 2 suggested that the teachers seat the children in the group so that they were facing the teacher and the teacher was facing the rest of the class. This was to serve the dual purpose of maintaining the attention of students in the group once the lesson began, and also of improving the teacher's ability to monitor the classroom.

Implementation of this principle was measured with a five-point rating scale in which the observer noted the appropriateness of both the teacher's seating and the children's seating as the percent of the rest of the class which could be seen by each. There were no differences in the average rating of appropriateness of teacher seating (5301), with the control teachers averaging 3.30, and the treatment teachers averaging 3.63 (with 3 representing an unobstructed view of 60% of the class). However, there was a difference ($p = .09$) in the average rating of child seating (5302). The control teachers averaged 2.17 and the treatment teachers averaged 2.72 (with a rating of 2 indicating that students in the group could see 80% of the other students).

Summary of group differences for getting and maintaining the students' attention (Principles 1 & 2). It was expected that the treatment teachers would be more likely than the control teachers to use signals to get students' attention at the beginning of transitions and lessons, and that once the students were in the groups, treatment teachers would use seating arrangements to maintain student attention and minimize distraction from the rest of the room. The results indicate that these two principles were not implemented more by the treatment group.

There were few differences between the treatment and control groups regarding their use of signals to start transitions and begin lessons. Both groups often used signals for transitions, but neither group used signals to start lessons very frequently. There were some differences in the types of signals used, but these cannot be attributed directly to the model. There were some results indicating that the treatment teachers may have had more efficient transitions, although these were not uniformly strong. Also, this cannot be directly related to the principle, because it emphasized the use of a clear signal to achieve quick transitions and the groups did not differ here. However, it is possible that having their attention drawn to transitions may have made the treatment teachers more aware of how well they organized and conducted them.

Results for variables measuring group seating also did not indicate high implementation in either group, although they suggested that the treatment teachers were slightly more likely to position their reading group students appropriately according to Principle 2 (i.e., their backs to other students in the room).

Therefore, there was some evidence of a treatment effect for these two principles, but it was not very strong. The teachers in both groups were already "implementing" both principles to some extent, and the treatment did not increase this level of behavior significantly.

Introducing the Lesson and New Material to the Students (Principles 3,4,5, & 6)

Using an overview. Principle 3 suggested that the teacher should introduce each lesson with a brief overview. Implementation of this principle was measured by noting what kind of overview of instructional content was given during each lesson and when it was given. The observers noted whether an overview of the instructional content was given, an overview of the mechanics of the lesson was given ("We are going to read four pages today."), or no overview of any kind was given. They also noted the inclusion of motivational statements in an overview, and whether these were specific or nonspecific. They also rated the enthusiasm of the teacher's voice during the overview, and the apparent effect on students' enthusiasm. It was expected that the treatment teachers would be more likely to use overview presenting specific content.

There was no significant difference between the groups as to whether or not an overview was given (5330), although there was a trend (p = .11) in the expected direction, with the control teachers giving no overview 66% of the time while the treatment teachers failed to give one 51% of the time. This suggests that there might have been some treatment effect, although the principle was far from being implemented on a regular basis by the treatment teachers.

When an overview was given, there was no difference between the two groups for inclusion of motivational statements (5334). There was a difference (p = .06) for the average rating of enthusiasm in the teacher's voice (5305), with the control teachers being rated slightly higher on a five-point scale (2.42 vs. 2.20). There were no differences in the effect of overviews on students' enthusiasm (5306).

It must be concluded that Principle 3 was not implemented on the basis of the treatment.

Presenting new words. Principle 4 stated that new words should be presented before the students encountered them in the lesson, so it was expected that treatment teachers would have a higher proportion of new words given at the beginning of the lesson rather than during it. The teacher's presentation of new words was measured by noting when they occurred in the lesson, whether they were given at the beginning or the end of the lesson, whether they were given by the teacher or asked of a child, and what kind of clues were given when the words were presented.

The expected difference was not found. Both groups presented the majority of words at the beginning of the lesson (5359; means = .72 for control and .73 for treatment). For this principle therefore, there was "implementation" independent of the treatment (as demonstrated by the control group's score), and the treatment did not increase this behavior over the base rate of the control group.

There was a significant effect ($p = .09$) for the proportion of new words which were given by the teacher rather than being asked of a child (5360). The control teachers gave the words 42% of the time, while treatment teachers only did this 20% of the time. This behavior was not specified in the treatment, but might be a reflection of other principles which emphasized the importance of student responses.

There was also a significant difference between the treatment groups in the types of clues given whenever new words were presented (5361-5364). The treatment group was more likely ($p = .10$) to present only phonetic clues whenever new words were given to the students (34% and 54%). Again, this behavior was not specified by the treatment, so the effect cannot be related to it directly.

There were treatment effects for two of three variables indicating the number of presentations of new words. In both cases, the treatment group was higher than the control group. The average number of new words presented during a lesson (total number of new words over the year divided by the number of observation) was different (5358; $p = .07$; means = 1.00 for control group, 1.93 for treatment group). The average number of new words which were presented at any one time was also higher for the treatment group (5369; $p = .03$; means = 3.92 for control group, 5.48 for treatment group). There were no significant differences in the proportion of lessons in which *any* new words were presented, although the direction of the differences

CI

was the same (5368; means = .24 and .34).

Repetition of new words. In order to measure the implementation of Principle 5, repetition of new words, the observer noted for each new word whether or not it was repeated by the students and how this was done. There were no significant differences in the proportion of new words that were repeated by the students, and the trend was not in the expected direction (5365; means = .47 and .27). There were also no significant differences in the proportion of new words that were repeated by all of the students rather than by only some of the students (5367; means = .75 and .73). The first of these measures suggests that there was fairly low implementation of this principle for the treatment group and higher use in the control group, although this difference was not significant. However, each group usually had all of the children repeat when repetition was used. There was a significant difference ($p = .02$) in the proportion of new words that were repeated by choral repetition rather than individual repetition (5366; means = .83 and .47). It may be that Principle 7, which discouraged choral responses, caused treatment teachers to avoid having students repeat new words (choral responses may have been the most typical way of getting this done efficiently).

Demonstrations and explanations. In order to measure implementation of Principle 6, teachers' use of demonstrations and explanations to precede student activities, the observer noted whenever a demonstration occurred, rated the sufficiency of that demonstration in his or her opinion, rated the students' apparent comprehension of the demonstration, and noted the ways in which the teacher checked student comprehension. When activities were begun without demonstrations, the observer noted whether or not they should have been explained (based on inferences about student confusion), and also whether or not a demonstration was repeated because students did not understand. It was expected that the treatment teachers would have more effective demonstrations, and therefore higher ratings for sufficiency and student comprehension, and fewer repetitions of demonstrations. It was also expected that they would have a higher proportion of activities introduced by teacher demonstration when necessary, and that they would be more likely to deliberately check for student comprehension before beginning activities.

None of these hypotheses was supported by the data, and there was an unexpected result for the last one. There were no significant differences for the average rating of sufficiency of demonstration (5310). On a five-point scale, the control mean was 2.66 and the treatment mean was 2.68. There was no difference in ratings of student comprehension of demonstrations (5311; means = 3.74 for control and 3.62 for treatment). There also were no differences in the proportion of demonstrations that had to be repeated due to student confusion (5351; means = .02 for control and .04 for treatment).

There also were no differences in the proportion of activities introduced by teacher demonstration (5348; means = .93 for control, .91 for treatment). The means for this variable indicate a high level of implementation in the absence of treatment, with no treatment effect above that base rate.

There were five categories for coding teacher checking of students' comprehension of the demonstration: asking them questions, having them repeat the instructions, having them demonstrate the procedure in front of the group, starting the lesson while in the group and observing students' ability to complete the activity, and sending students back to their seats after the demonstration without checking for comprehension. It was expected that the treatment teachers would be less likely to dismiss students without checking, but this was not the case. In fact, the control teachers were less likely (5357; p = .03) to do so, with 27% of their demonstrations ending this way, compared to 40% for the treatment group. The control teachers were more likely (p = .01) to keep the students in the group after a demonstration and observe them while they did the activity (5356; means = .38 and .17). There were no significant differences in the other three ways of checking comprehension: asking questions (5353; means = .21 for control group and .26 for treatment), having students repeat instructions (5354; .02 and .03) and having the students demonstrate procedures before dismissal to do the work (5355; .16 and .19). Therefore, teachers in both groups were doing something to check comprehension most of the time, but the treatment teachers were doing this less often than the control teachers.

Summary of group differences for introducing the lesson and new material to the students (Principles 3,4,5, & 6). The four principles in this part of the model were not implemented by the treatment group more than the control group. In fact, the few differences that did exist either could not be related directly to the instructional model or showed unexpected results. It had been expected that the treatment group teachers would give overviews more frequently, would present new words at the beginning of the lesson more often, would have students repeat them more often, and would give more and better demonstrations, checking to make sure that students understood them.

There were no differences for the use of overviews, and neither group used them more than half the time. (There was a nonsignificant trend in the expected direction, however.) The variables describing the introduction of new words did not reveal any differences that could be attributed directly to the treatment. (However, treatment teachers presented more new words, and were more likely to use phonetic clues when presenting them). Most teachers in both groups presented new words at the beginning of the lesson, as suggested in the treatment. There were also no differences in teachers' use of demonstrations to precede new tasks, as teachers in both groups did this a great deal. The control teachers were more likely to check students' comprehension before dismissing them to their seats to do assignments, which was not expected.

Therefore, it must be concluded that the instructional model did not influence the treatment teachers to systematically introduce the lesson and new material using the methods described in Principles 3 through 6.

Calling on Individual Students in the Group (Principles 7,8,9,10,11, & 12)

Working with one child at a time and giving feedback. Principle 7 suggested that the teacher work with one child at a time, offering opportunities to practice new skills and receive feedback. An implication of this was that choral responses should be minimized. The implementation of this principle was measured by three types of variables: the rate of individual response opportunities (R.O.'s), the proportion of response opportunities which received no feedback from the teacher, and the use of choral or group responses. It was expected that the treatment teachers would have a higher rate of R.O.'s and lower rates of "no feedback" and choral or group responses.

Response opportunities were analyzed for total R.O.'s summed over reading turns and nonturn interactions, as well as separately for each of these situations. (Refer to the discussion of the observation system and creation of variables in Chapter 1 for further explanation of this procedure.) The rate of total R.O.'s was different for the two groups (601; *p* = .07, control mean = 1.86 interactions per minute, treatment mean = 2.20 per minute). There was also a significant difference in the number of nonturn interactions per minute (602; *p* = .02; control mean = 1.20, treatment mean = 1.64). There were no significant differences in the number of reading turns per minute offered to the students (603; control mean = .40 and treatment mean = .36), or in the number of reading turn interactions per minute (604; control mean = .65, treatment mean = .56).

A variable related to this was the proportion of all response opportunities which occurred during reading turns. Here, there was a significant difference between the two groups (605; *p* = .05, means = .36 and .26). This suggests that the treatment teachers asked proportionately more single questions, so that more interactions occurred outside of oral reading turns.

There were no differences between the groups in the absolute rate per minute of no feedback to R.O.'s (24; means = .06 for control and .10 for treatment). There were also no differences in the proportion of total R.O.'s which received no feedback (723; means = .07 for control and .10 for treatment). There was a difference (*p* = .10) in the unexpected direction for the proportion of nonturn R.O.'s which received no feedback (724; means = .04 for control and .09 for treatment). These results suggest that teachers in both groups gave feedback to individual students most of the time, and therefore were actually implementing this principle naturalistically. However, treatment did not produce higher implementation.

When response opportunities were broken down by type of answer, it was evident that neither group of teachers responded to incorrect answers very often with no feedback (793, 794; means for total incorrect answers were 0 for both groups, and means for nonturn incorrect answers were both .01). However, responding to correct answers with no feedback was more likely for the treatment group than the control group (778: for total correct R.O.'s, $p = .10$, with means = .04 for control group and .10 for treatment group; 779: for nonturn correct R.O.'s, $p = .09$, with means = .05 for control and .11 for treatment). ("No feedback" was not examined separately for turn interactions because it happened so seldom there.)

There were fairly strong treatment differences in the expected direction for the use of group responses rather than individual responses. These were measured by the observer counting the frequency of teachers' use of choral responses (the teacher indicated to the students that she wanted them to respond together) and the use of group call outs (more than one student shouted out the answer to a question together and this response was accepted by the teacher, even though she had not indicated that she wanted a group response). These frequencies were analyzed as rates per observation and as rates per minute. When considered as the average number per observation, there were no significant differences between the groups for choral responses alone, although the trend was in the expected direction (4038; means = 3.92 for control and 2.38 for treatment). There were differences for the average number of group call outs (4039; $p = .08$; means = 6.10 for control and 3.27 for treatment) and for the average total of choral responses and group call outs (4040; p = .05; means = 10.02 for control and 5.65 for treatment). When these same three frequencies were examined in terms of rate per minute of observation, all three were significant: the rate of choral responses (4051; p = .10, .17 for control and .10 for treatment), the rate of group call outs (4052; p = .03, .27 for control, .13 for treatment), and the rate of total choral responses plus groups call outs (4053; p = .01, .44 for control and .23 for treatment). These results indicate that even the treatment teachers were using group responses some of the time, but their level of use was much less than that of the control group. Therefore, it can be concluded that there was a treatment effect for minimizing group responses.

Types of initial selection of respondents. To measure implementation of Principles 8, 10, and 11, observers noted for each R.O. the type of selection used to call on the students initially. Initial selection meant the way the student was selected for the first interaction within a possible sequence of interactions. Therefore, this was distinguished from interactions which were the result of sustaining feedback in the immediately preceding interaction. The five types of initial selection were: ordered turns (the teacher chose the student on the basis of seating position), preselection (the teacher named the student and then asked the question, and was not relying on seating order to choose the child), nonvolunteer (the teacher asked the question, and then called on a student who had not raised his or her hand and the teacher was not relying on seating order to make the selection), volunteer (the teacher called on a student who had raised his or her hand, and who was not selected because of seating order), and call out (the teacher asked a question and some student called out the answer without first receiving permission, and the content of this answer was responded to by the teacher). It was expected that teachers in the treatment group would be more likely to use ordered turns, and less likely to use volunteer selections, except for personal questions, and would have lower rates of calling out by students.

When absolute rates of these selection types were examined, differences in the expected direction were evident. There were significant differences for three of the rates: ordered turns (1; p = .05, means = .22 for control and .56 for treatment), nonvolunteer (3; p = .05, .17 for control and .08 for treatment), and volunteer (4; p = .07, .12 for control and .06 for treatment). There were no significant results for the absolute rate of preselections (2; means = .08 for control and .05 for treatment) or for call outs (5; means = .07 for control and .04 for treatment).

Another way to examine implementation of these principles is to look at the proportion of response opportunities involving initial selection by each of these methods. These were examined separately for total interactions, selection of reading turns, and nonturn interactions. They were also broken down as to type of question: reading questions, in which the student was asked to decode words or answer questions about decoding; nonreading questions, which were related to the reading lesson but which did not involve actual decoding skills; and personal questions, which required an opinion or personal experience of the student.

The first set of variables concerns general response opportunities (without breaking them down by question type). The use of ordered selection was subject to a strong treatment effect as evidenced by the proportion of initial selections which were ordered: for total interactions

69

(607; p < .01, means = .24 for control and .72 for treatment), for selection of students to begin reading turns (608; p < .01, means = .29 for control and .82 for treatment), and for selection for nonturn interactions (609; p < ,01, means = .22 for control and .69 for treatment). It is evident from these results that teachers in the treatment group were using ordered selection the majority of the time to select respondents, as was suggested in the treatment. The control teachers were using it to some extent, but not as much as the treatment teachers.

Preselection was not used very often by either group, but it was used significantly more often by the control group as measured by the proportion of initial selections that were preselections: for total interactions (610; p = .01, means = .15 for control and .04 for treatment), for selection for reading turns (611; p = .01, means = .19 for control and .05 for treatment), and for selection for nonturn interactions (612; p = .03, means = .12 for control and .04 for treatment).

There were also strong treatment effects for the use of nonvolunteer selection, with the control teachers using this technique significantly more than the treatment teachers. Statistics for the proportion of initial selections which were nonvolunteer were: for total interactions (613; p < .01, means = .31 for control and .11 for treatment), for selection for reading turns (614; p < .01, means = .34 for control and .09 for treatment), and for nonturn interactions (615; p < .01, means = .29 for control and .11 for treatment).

Likewise, the proportion of initial selections which were volunteers was greater for the control teachers than for treatment teachers: for total interactions (616; p = .01, means = .19 for control and .08 for treatment), for selections for reading turns (617; $p = .01$, means = .17 for control and .03 for treatment), and for nonturn interactions (618; $p = .03$, means = .20 for control and .10 for treatment).

Significant differences were also present for the proportion of initial selections which were *call outs*: for total R.O.'s (619; $p = .01$, means = .11 for control and .05 for treatment), and for selection for nonturn interactions (620; $p = .01$, means = .17 for control and .06 for treatment). Call outs were not used to select students for reading turns often enough for separate analysis.

When initial selections were broken down by type of question, the same patterns were evident, although they were not as strong for some types of questions. Again, the proportions of interactions selected by ordered turns showed highly significant differences. The p-levels for these comparisons were .01 for all types of questions: total reading questions (621), nonturn reading questions (622), total personal questions (623), and total nonreading questions (624). In all cases the treatment teachers had higher means.

There were also significant differences in the proportion of initial selections made by preselections: for total reading questions (625; $p = .01$), for nonturn reading questions (626; $p = .02$), and for total nonreading questions (627; $p = .06$). In all cases, the control teachers used preselections more.

The pattern was similar for use of nonvolunteer selections. There were significant differences for total and nonturn reading questions (628, 629; $p < .01$ for both), for personal questions (630; $p = .05$), and for nonreading questions (631; $p = .01$), with the control teachers using nonvolunteer selections more for each type of question.

In the treatment, it was suggested to the teachers that volunteering should be minimized overall, especially for academic questions, but that it might be used appropriately for personal questions. This suggestion was apparently understood and implemented, because the proportion of initial selections which were volunteers shows group differences for total and, for nonturn reading questions (632, 633; $\underline{p} < .01$ for each), but not for personal questions (634; $\underline{p} = .19$). There were also no differences in the proportion of initial nonreading questions selected by volunteering (635). If it can be assumed that reading questions are more likely to require students to practice important skills, these results indicate that the treatment teachers were being careful to minimize volunteering in this more important situation, but were allowing it under other circumstances.

Call outs also were examined by type of question. There were no differences in the proportion of initial nonturn reading questions selected by call outs (637), or the proportion of initial personal questions selected by call outs (638). There were significant differences in the proportion of initial nonreading questions (639; $\underline{p} = .01$), and in the proportion of initial total reading questions (636; $\underline{p} = .01$), which were selected by call outs. In both cases, the control teachers had more call outs.

Those results indicate high implementation by the treatment group of the recommendations for selection techniques. Ordered selection was used only about 30 percent of the time in the absence of treatment, but this was more than doubled in the treatment group. The control group, instead, tended to rely primarily on nonvolunteering and volunteering selections with some use of preselections and some allowance of call outs. Therefore, they allowed more student control of selection through volunteering and call outs, and when controlling selection themselves, typically they did not do it in a systematic fashion.

Principle 11 suggested that teachers discourage call outs, and also made some suggestions for ways of dealing with them when they occurred. Implementation of this part of the principle was measured by the observers noting when call outs were accepted for academic content, when they were also corrected, when they were not accepted but were corrected, and how they were corrected. It was expected that the treatment teachers more often would correct call outs (remind the student not to call out) and refuse to accept their content by responding to the answer itself. The proportion of call outs which were accepted but which were also corrected by the teachers was higher in the treatment group than the control group (1032; $\underline{p}$ = .01, means = .01 for control and .07 for treatment). However, this significant difference does not equate to high implementation, because the treatment teachers did not correct most call outs when they did occur, as was suggested by the treatment. There was no significant difference in the proportion of corrected call outs that were not accepted for their content (1001; means = .93 for control and .87 for treatment). These data indicate that when the teachers in either group did stop to correct call outs, they were not likely to also accept the content of that call out.

Four categories of corrections of call outs were examined: management statements (mild corrective statements), warning (moderately severe corrective statements), criticism (severe corrective statements), and ignore. When all corrected call outs were considered, there were no differences in the ways that control teachers and treatment teachers dealt with them

1037-1039). When only unaccepted call outs were examined for type of correction, again there were no group differences in the ways teachers dealt with them (1020, 1023).

For both unaccepted and accepted call outs, teachers were most likely to use management statements to correct the students, rather than more severe statements (1037; means = .53 of total call outs corrected with management for control and .61 for treatment). One of the suggestions in the treatment was that call outs should be corrected fairly mildly (so as not to discourage enthusiasm but to teach the children to channel it in a different direction). These data suggest that most teachers were using fairly mild corrections most of the time anyway, and the treatment did not significantly increase this type of behavior for the treatment group.

Use of comments. Principle 9 suggested that the teacher could use occasional out-of-turn comments in order to maintain students' attention. However, this suggestion was not implemented by the treatment group, and there was little or no natural implementation in the control group. The means rounded to two places for the absolute rates per minute were .00 for both groups, although there was a significant difference at a level of $p = .02$ (8). The proportion of nonturn response opportunities which were comments was not significantly different, although this bordered on significance (665; $p = .11$, means = .00 for both groups). The F ratios and p values for these data suggest that there were group differences, but the means were so low that the direction is not detectable at two decimal places, and therefore, the results are meaningless.

Use of undesirable questions. Principle 12 suggested that teachers avoid the use of questions which might be confusing to students. In order

74

to measure implementation of this principle, observers noted the frequency of the following categories of "undesirable" questions: rhetorical, answering of one's own question without waiting for students to answer, repeating questions, other types of confusing questions, and total number of undesirable questions. These were examined in terms of the average number per observation, as well as the average rate per minute. Many of these variables were not suitable for analysis because of low frequency, and those which were examined yielded no significant results (4045-4049, 4055, 4058). Undesirable types of questions did not occur often (the average total per observation was only .70 for the control group and .56 for the treatment group). Therefore, there was natural implementation of this principle, even in the control group, because the problem did not occur very often, and no effect on the treatment group beyond the base rates was present.

Summary of results for principles relating to calling on individual students in the group (Principles 7,8,9,10,11, & 12). This set of principles made suggestions about giving response opportunities to individual students in the group, providing feedback to response opportunities, and selecting students for R.O.'s, along with general guidelines for types of questions to avoid. It was expected that treatment teachers would have: a lower rate of group responses; a higher rate of individual responses; a lower rate of "no feedback" from the teacher; a higher rate of ordered selection and a correspondingly lower rate of volunteer selections and call outs, at least for academic questions; a higher rate of use of comments by other students; and a lower rate of use of undesirable questions.

Many of these expectations were supported. Treatment teachers did exhibit a lower rate of group responses, especially group call outs, and a higher rate of individual response opportunities, especially in nonturn interactions (single questions occurring outside of oral reading turns).

There were unexpected findings for the use of "no feedback" to answers, in that the treatment teachers were more likely to fail to give feedback than the control teachers, but only in response to correct answers, when failure to give feedback is a less serious problem.

Treatment teachers had a much higher rate of use of ordered selection, in which respondents were called on in order around the group. Correspondingly, treatment teachers were less likely to rely on other types of selection: volunteering, preselection, nonvolunteers, and call outs. These results were strongest for reading questions.

There were no clear differences between the two groups in their use of student comments (there were significant differences, but the means showed very low levels of use in either group, and the direction is not evident). There also were no differences between the groups in the use of undesirable questions, with both groups demonstrating very low levels of use.

In general, the principles in this section were implemented to different degrees, with the strongest effects being on the use of individual response opportunities and ordered selection. In fact, the results for use of ordered turns probably were the strongest for any single principle. Apparently, the suggestions given about this in the materials were sufficiently clear and reasonable to encourage the treatment teachers to try the technique, and their high implementation of it suggests that it was useful to them.

Dealing with Individual Learning Rates within the Group (Principles 13,14,15, & 16)

Breaking up the group. Principles 13 and 14 were concerned with breaking up the reading group whenever a few students were not meeting the lesson's objectives. It was suggested that one way to accomplish this was by teaching the more able students through to a certain point and dismissing them, in order to spend more time concentrating on those few students who were not meeting objectives. It was expected that the treatment teachers would do this more often than the control teachers.

To measure implementation of this principle, observers noted for every observation of a group whether or not the group was broken up due to ability differences in this way, whether it needed to be broken up but was not, or whether it did not need to be broken up. The observers made these distinctions on the basis of whether one or two students in the group were notably behind that day, in their judgment. The observers also noted certain things about the way a group was broken up when it did occur.

There were no significant differences between the treatment and control groups for the proportion of observations in which the groups were broken up, needed to be broken up, or did not need to be broken (5338-5340). The means for these three variables, respectively, for the control group were .01, .06, and .92, and for the treatment group were .03, .07, and .89. These results suggest that there was low implementation of this principle in the absence of treatment, and that the treatment did not significantly increase the teachers' use of the technique. The results also suggest that there was less need for breaking up the group due to ability differences than was expected beforehand. The descriptions of the various ways in which groups were broken up were not analyzed, because this occurred so few times.

Using other students as models. Principle 15 suggested to the teachers that they use other students as models on occasion to help those students who were having problems. It was therefore expected that treatment teachers would do this more than control teachers. Implementation of this principle was to be measured by noting when it occurred, and how it was used when it did. However, no teachers used students as models at any time throughout the year. Obviously, this result represents failure to implement the treatment, and it demonstrates that this technique was not used by teachers in the absence of the treatment. Perhaps the content of first-grade reading instruction is such that models are not as useful as they might be in other curriculum areas, or perhaps this technique is difficult to use well, and the brief discussion in the treatment materials was not sufficient to encourage its use.

Providing extracurricular help. Principle 16 recommended extra tutorial help for students not meeting lesson objectives. Since this could not be measured directly with the observational system, pertinent data are not reported in this paper. Data will be presented and discussed in a later paper concerned with the teacher interviews conducted at the end of the school year.

Summary of principles related to dealing with individual learning rates within the group (Principles 13,14,15, & 16). Principles 13, 14, and 15 had worse implementation than any other group of principles. Although it had not been expected that the treatment teachers would use the suggested techniques on a daily basis, it was expected that they would use them more often than the control group teachers. However, there were very few instances of breaking up the group because of ability differences in the way suggested, and there were no instances of use of models in either control or treatment classrooms.

One possible conclusion is that the behaviors suggested by these principles are inappropriate for first-grade reading groups, so that there was no implementation because the teachers judged them as such. Another possible conclusion is that the techniques might be useful, but the minimal treatment was not sufficient. These principles were different from many of the others in that they were asking the teachers to try something novel. Most of the other principles were asking the teacher to use behaviors already familiar to them, but perhaps more systematically. It seems likely that complex or novel behaviors will require more extensive treatment. The more a treatment or a program requires a teacher to change from his or her normal repertoire of teaching behaviors, the more necessary it will be to provide extensive rationales and opportunities for practice and feedback. These were not provided in this study, and the data suggest that they were needed for the principles discussed in this section.

Responding to Answers that are not Correct (Principles 17,18, & 19)

Use of terminal and sustaining feedback. Principles 17, 18, and 19 were concerned with the responses teachers gave to answers that were not correct, including failures to respond.

Implementation of these principles was measured with the low-inference part of the observation system. For every academic interaction between the teacher and the individual student, the observer noted the type of selection, the type of question asked, the quality of the answer received (whether correct, incorrect, don't know, or no response), and the type of feedback given by the teacher to that answer. The catagories of feedback which were related to these principles were terminal feedback, in which a student's opportunity to improve an incorrect answer was terminated and the answer was given, either by the teacher or another child, and sustaining feedback, in which the teacher stayed with the student who had made the mistake, and continued to ask questions or to probe in an attempt to help the student improve his or her answer during another interaction.

Three categories of terminal feedback were used: the teacher could give the answer to the student herself; she could ask another student to give the answer to the child; or another student could call out the answer before the teacher had a chance to give feedback. Sustaining feedback also was coded into three categories: the teacher could simply repeat the question to the student without providing additional information; she could give clues to the student in the form of simplifying questions which required another response; or she could essentially give the answer by a clue (by asking a very simple question with an obvious answer that allowed the student to make a correct response).

In addition to noting type of sustaining feedback, the observer would note whether or not each use of sustaining feedback led to an improved response by the student. Therefore, for each interaction involving one of the three types of sustaining feedback, the observer noted whether it led to an improved response or an unimproved response. Improvement was considered to be any response following a "no response," or some correct answer following an incorrect answer.

It was expected that the treatment teachers would use more sustaining feedback than the control teachers. It was also expected that the treatment teachers would give the answer to the students on occasion, when clues were not appropriate, but would have a lower rate of use of this technique than the control teachers, because clues or other sustaining feedback were suitable for most questions. It was also expected that the treatment teachers would have a lower rate of asking another student for the answer.

As with other variables from the low-inference observation system, results are reported in terms of absolute rates per unit of time, and relative rates (proportion of possible times a particular behavior occurred). Variables are discussed separately for total interactions, for those occurring only in reading turns, and for those occurring outside of reading turns.

There were no significant differences between the groups in the absolute rate of use of sustaining feedback (7). However, there was a significant difference ($p = .06$) for the proportion of all interactions which were the result of sustaining feedback (648; means = .12 for control, .15 for treatment). Separate analysis for nonturn interactions yielded a significant difference between the two groups for the proportion of interactions that were the result of sustaining feedback (652; $p = .02$, means = .11 for control and .14 for treatment).

For any noncorrect answer (incorrect, "don't know," or failure to respond), the teacher had to use either terminal or sustaining feedback. Therefore, a score was computed for each teacher for the proportion of these interactions that involved terminal feedback rather than sustaining feedback. That is, when given a choice, which of the two strategies did the teacher use more often? There were significant differences in use of terminal feedback for total response opportunities (757; $p < .01$, means = .62 for control and .44 for treatment), but not for interactions occurring during reading turns (758; means = .66 for control and .57 for treatment). There was a significant difference for nonturn interactions (759; $p < .01$, means = .52 for control and .36 for treatment). These data suggest that the treatment teachers were more likely to use sustaining feedback in situations where it was an option, especially in nonturn interactions, although they did not use it all of the time. Their proportionate rate of use was significantly higher than that of the control group, although the control group also had some natural implementation in the absence of treatment. Therefore, the result of the treatment was to increase the use of this behavior over the base rate of natural implementation.

To examine use of feedback more specifically, variables were computed to indicate the extent of use of each of these separate types of feedback within the two major headings. When absolute rates per minute were examined, there were no significant differences for the use of any category of terminal feedback or for the first two categories of sustaining feedback (30-32, 34, 35). There was a difference in the absolute rate for the third type of sustaining feedback, *giving the answer by a clue* (36; $p = .07$, means = .00 for control and .01 for treatment). These means certainly do not indicate high implementation by the treatment group, but they do indicate that treatment teachers were using this behavior some of the time, while control teachers hardly ever or never used this particular technique.

Variables were computed which examined the proportion of response opportunities which involved each of the separate types of feedback. When the proportion of total response opportunities which included *give answer* feedback was compared, there was a difference between the two groups in the expected direction (735; $p = .07$, means = .15 for control and .09 for treatment). When the data were broken down into turn and nonturn interactions, there were no significant differences (736, 737). However, the means indicate that giving the answer occurred much more frequently in turn response opportunities than nonturn. Means for the two groups for the proportion of turn response opportunities which included giving the answer (736) were .49 for control and .41 for treatment teachers. Means for nonturn interactions (737) were .03 for control and .04 for the treatment group.

When the two groups were compared for the proportion of total response opportunities which included *asking another child for the answer*, there was a significant difference in the expected direction (738; $p = .06$, means = .05 for control and .03 for treatment). This variable was not analyzed for reading turn interactions, but a significant difference occurred with nonturn interactions (739; $p = .04$, means = .07 for control and .03 for treatment). Even though a significant treatment effect is present, the means indicate that the proportion of interactions in which the teacher did ask another student was very low. This technique was not used extensively by these teachers in their reading groups even in the control group, although the treatment did seem to reduce this naturally low level.

There was a significant difference for the proportion of total response opportunities which involved students calling out feedback, and this was in the expected direction (740; $p = .06$, means = .02 for control and .01 for treatment). When analyzed separately for turn interactions, the difference was not significant (741). When examined only for nonturn interactions, there was a significant difference, but the direction was not detectable (742; $p = .03$, means = .01 for control and .01 for treatment). The same conclusion can be drawn for the use of *call out* feedback as for *ask other* feedback. That is, there was a low incidence, even in the control group. However, the treatment perhaps did serve to reduce this slightly.

The proportion of response opportunities which involved the teacher repeating the question was not significantly different across groups (746, 747, 748; means = .05 for both groups for total interactions). This type of sustaining feedback was used infrequently in the control group, and its rate was not affected by the treatment.

There were treatment effects for the other two types of sustaining feedback. The proportion of all response opportunities that involved clue feedback was significantly different for the two groups in the expected direction for total R.O.'s (749; $p < .01$, means = .06 for control and .09 for treatment). The same variable was also significant for turn interactions (750; $p = .03$, means = .13 for control and .19 for treatment), and nonturn interactions (751; $p = .02$, means = .06 for control and .09 for treatment). These results suggest that the treatment did increase the use of this feedback technique a small but significant amount.

The last variable of this type involved the proportion of response opportunities involving give by clue feedback. There were significant differences for total interactions for this variable (752; $p = .02$, means = .00 for control and .01 for treatment), and for nonturn interactions (754; $p < .01$, means = .00 and .01). There were no significant differences for this variable for turn interactions (753; means = .01 and .02). These results suggest that this technique was not used much in the absence of treatment. The effect of the treatment was not overwhelming, but it did increase the teachers' use slightly.

The proportions given for the above variables are low because the

84

denominator in this case is all response opportunities offered for either total, turn, or nonturn settings. This includes correct answers, which would not usually be occasions for these kinds of feedback. Therefore, the same types of variables were created for the specific types of answers: incorrect, "don't know," and no response. These are examined in a later section.

Because all teachers, both control and treatment, used both terminal and sustaining feedback some of the time, each of these major categories was examined for the specific type of feedback used. It was expected that there would be times when terminal feedback would be the most sensible thing for the teacher to do, either because of the pace of the lesson or because of the type of question. However, when taking this option, it was expected that treatment teachers would be less likely to ask other students or to have call outs, and therefore probably would have proportionately more use of giving the answer.

When the proportion of all terminal feedback which was giving the answer by the teacher was examined, there were no significant differences for total or for reading turn response opportunities (760, 761; means for total R.O.'s were .62 and .72, and means for turn R.O.'s were .85 and .93 for control and treatment groups, respectively). However, when only nonturn response opportunities were examined, there was a significant difference between the two groups (762; $p < .01$, means = .29 and .51 for control and treatment).

When the proportion of all terminal feedback which was asking other students for the answer was examined, there was no significant difference for total response opportunities (763; means = .27 and .20). There was a significant difference for nonturn interactions (764; $p = .01$, means = .57 for control and .39 for treatment). This variable was not examined for reading turn interactions alone.

The proportion of terminal feedback which was a *call out by another student* before the teacher could respond was not different for the treatment groups for either total interactions (765; means = .11 and .08), turn interactions (766; means = .12 and .07), or nonturn interactions (767; means = .14 and .10).

These data on types of terminal feedback suggest that the treatment group was relying more on the teacher giving the answer herself than on asking or allowing another student to do so, while the control group was more likely to ask another student, especially in nonturn interactions. This represents a treatment effect, although it is evident that both groups of teachers used all techniques some of the time. The treatment effect therefore was to modify the frequency with which each technique was used, not whether or not it was used.

The three types of sustaining feedback were examined in the same way. There were no significant differences for any type of interaction for the proportion of sustaining feedback which was *repeating the question*. The means for total interactions, for the control and the treatment groups respectively, were .41 and .32 (768). For turn interactions they were .42 and .35 (769), and for nonturn interactions they were .41 and .33 (770).

There were also no significant differences in the proportion of time that *clue* feedback was used to sustain an interaction. (771, 772, 773; means for total interactions were .55 and .61, for turn interactions .53 and .60, and for nonturn interactions .55 and .60).

There were significant differences in the proportion of all sustaining feedback which was give by clue feedback, for total interactions (774; $p = .07$, means = .04 and .07). This was not significant for turn interactions (775; means = .04 and .05), but it was for nonturn interactions (776; $p = .01$, means = .03 and .07). These results suggest that both treatment and control teachers used the techniques of repeating the question or offering clues to the students about the same proportion of time that sustaining feedback was used. (Remember, however, that the rate of use of sustaining feedback in general was different for the two groups.) However, very simple clues that essentially gave the answer to the student but which allowed him or her another response opportunity were seldom used by either group, but more often by the treatment teachers. Possibly this technique is not one that is readily apparent to teachers, and discussing it in the treatment made many treatment teachers aware of it. However, it was still not used by the treatment teachers as much as other, more typical types of sustaining feedback.

Feedback to "no response" answers. To examine the use of each feedback technique more closely, variables were created which examined particular subsets of response opportunities defined by the type of answer involved: no response, incorrect, or "don't know." The denominator of the proportion therefore becomes, for example, "all no response answers" rather than "all response opportunities."

There were no significant differences between the two treatment groups on the absolute rate of no response answers per minute (23; means = .15 for control and .12 for treatment). However, there was a difference between the two groups for the proportion of total response opportunities which were no response answers (698; p = .06, means = .14 for control and .10 for treatment group). When this variable was broken down into turn and nonturn interactions, there were no significant differences (699, 700).

The proportion of no response answers which were responded to by the teacher giving the answer to the student was not different for the two groups for total interactions (833; means = .42 and .32) or for turn interactions (834; means = .61 and .53). However, there was a significant difference for the use of this technique following no response in nonturn interactions (835; p = .10, means = .12 and .18).

When the use of ask other feedback in response to no response answers was examined, there was a significant difference for total interactions (836; p = .07, means = .16 and .08 for control and treatment respectively), and also for nonturn interactions (837; p < .01, means = .31 and .12). This variable was not examined separately for turn interactions.

The proportion of no response answers which involved another student calling out feedback showed no significant differences for total interactions (838; means = .09 and .05) or for turn interactions (839; means = .10 and .04). However, there was a significant difference for nonturn interactions (840; p = .06, means = .11 and .05).

There were no differences for the proportion of no response answers which received feedback in the form of repeating the question (844, 845, 846; means for total interactions were .10 and .12 for turn interactions they were .07 for both groups; and for nonturn interactions they were .14 and .16 for control and treatment, respectively).

However, there were significant differences for the proportion of no response answers which were followed by clues in total interactions

(847; $p < .01$, means = .21 and .37), turn interactions (848; $p = .02$, means = .18 and .32), and nonturn interactions (849; $p = .04$, means = .27 and .40).

Likewise, when the proportion of *no response* answers receiving *give-by-clue* feedback was examined, there were significant differences for all types of interactions: total (850; $p < .01$, means = .01 and .05), turn interactions (851; $p = .08$, means = .01 and .03), and nonturn interactions (852; $p = .01$, means = .02 and .06).

To summarize the data on teacher feedback to *no response* answers, the proportion of times teachers used either *clue* or *give-by-clue* feedback was significantly higher in the treatment group for all three types of interactions. The frequency of use of *asking another child* for the answer was significantly lower for the treatment group for nonturn, and total interactions. (This was not analyzed separately for turn interactions.) For nonturn interactions only, there were also differences between the two groups in the expected directions for giving the answer and having other students call out the answer when sustaining feedback was not used. That is, the treatment teachers used giving the answer much more often than either of the other two terminal feedback techniques, while the control teachers were more likely to have students call out the answer, or to ask another student to provide it.

During reading turn interactions, both treatment and control teachers were most likely to *give the answer* following *no response*. However, this was not true for nonturn interactions, when the most common response for the treatment teachers was to give a clue, and for the control teachers, to ask another child for the answer. Since the teachers were probably trying to maintain a faster pace in reading turns, it made sense for them to give the answer to the student then. However, even in the reading turn situation, both groups of teachers used clues or other types of sustaining feedback some of the time, although the treatment teachers were more likely to use it, especially clue feedback, than the control teachers.

These results can be interpreted as representing a fairly strong treatment effect characterized by an increase in the treatment teachers' use of sustaining feedback, especially clue feedback and give by clue feedback, and a decrease, especially for nonturn interactions, in their use of asking another student for the answer. These are the behaviors specified by the instructional model.

Feedback to incorrect answers. The same variables were examined for incorrect answers. In addition to looking at the six categories of feedback already discussed, the failure to give feedback (no feedback) and the use of emphasis feedback, were also examined for incorrect answers. No feedback was recorded by the observer when the teacher did not acknowledge the incorrectness of an answer and did not provide the correct answer. (This is related to variables discussed for Principle 7.) Emphasis was recorded whenever the teacher made a special effort to indicate what the answer was by repeating it. These two behaviors are relevant to examination of the principles because it was emphasized in the model that the teacher should communicate to the students when an answer was wrong.

There were no significant differences for the proportion of incorrect answers which were responded to with no feedback, for either total interactions (793; means for both were .00), or nonturn interactions (794; means for both = .01). This variable was not examined for turn interactions.

There were also no differences in the proportion of incorrect answers responded to with *emphasis* for either total interactions (795; means = .00) or nonturn interactions (796; means for both = .01). Obviously, neither group responded to incorrect answers very often with either of these types of feedback.

When the proportion of incorrect answers responded to by the teacher *giving the answer* was examined, there were no significant differences between the two groups for total interactions (803; means = .39 and .32), turn interactions (804; means = .55 and .52), or nonturn interactions (805; means = .16 and .18).

There were differences between the two groups for the proportion of incorrect answers responded to by the teacher *asking another child* for both total interactions (806; p = .04, means = .15 and .08), and nonturn interactions (807; p = .01, means = .28 and .14). This variable was not examined separately for turn interactions. These results are in the expected direction, with the treatment teachers *asking others* less often than the control teachers.

The proportion of incorrect answers responded to by other students calling out the answer showed no differences between the two groups for total interactions (808; means = .04 and .03), or for turn interactions (809; means = .04 and .03). However, there was a difference between the groups for this variable in nonturn interactions (810; p = .09, means = .04 and .02).

There were no differences between the two groups for the proportion of incorrect answers which were responded to by repeating the question for either total interactions (814; means = .21 and .22), turn interactions (815; means = .23 and .20), or nonturn interactions (816; means = .22 and .23).

However, there were significant differences for all types of interactions for the proportion of incorrect answers responded to with clues: total interactions (817; $p < .01$, means = .18 and .30), turn interactions (818; $p = .08$, means = .14 and .22), and nonturn interactions (819; $p = .02$, means = .24 and .35).

The proportion of incorrect answers responded to with giving the answer by a clue was not different for the two groups for total interactions (820; means = .01 and .03), or turn interactions (821; means = .02 for both groups). However, there was a significant difference for nonturn interactions (822; $p = .06$, means = .01 and .03).

In order to determine if types of teacher feedback to errors differed according to types of question, incorrect answers were analyzed for different types of feedback to reading and nonreading questions. These analyses revealed that the same pattern reported for incorrect answers as a whole was consistently found for both types of questions. That is, there was more use of sustaining feedback and less use of asking another student by the treatment teachers. (See variables 888-895, 899-907.)

To summarize the variables which describe teacher responses to incorrect answers, a pattern of results emerged which is similar to that for no response answers. That is, the treatment teachers had a higher relative use of clue feedback, and, in nonturn interactions only, give by clue feedback, suggesting implementation of the principles describing simplification. There also were differences between the two groups in the frequency with which they used asking another student for the answer, with the treatment teachers having a lower frequency of use than the control teachers. These results suggest that Principles 18 and 19 were often implemented by the treatment teachers when dealing with incorrect answers.

Feedback to "don't know" answers. "Don't know" answers were also examined for the types of feedback given to them. The proportion of "don't know" answers which were responded to by the teacher giving the answer was not different for the two groups for total interactions (823; means = .29 and .28), turn interactions (824; means = .60 and .56), or nonturn interactions (825; means = .12 and .18).

The proportions of don't know answers responded to by asking another child for the answer were not different for the two groups for total interactions (826; means = .17 and .12), but they were different for nonturn interactions, with control teachers using the technique more (827; p = .08, means = .27 and .16).

The use of other students calling out answers in response to don't know answers was not analyzed due to low frequency of occurrence.

The proportion of "don't know" answers responded to with the teacher repeating the question was not significantly different for either total interactions (828; means = .11 and .08), or nonturn interactions (829; means = .11 and .08). This variable was not examined for turn interactions.

The proportion of don't know answers responded to with a clue from the teacher was significantly different for the two groups for total interactions (830; p = .03, means = .23 and .36). This variable was not examined for turn interactions, but it was significantly different for nonturn interactions with treatment teachers using clues more (832; p = .05, means = .19 and .36).

The use of giving the answer by a clue as feedback to don't know answers was not examined because of the low frequency of such occurrences.

93

Although the results for don't know answers were not as strong as for no response and incorrect answers (possibly due to the lower frequency of "don't know" answers), similar patterns existed in that the treatment teachers were more likely to use sustaining feedback in the form of clues, and less likely to use asking another child for the answer. Again, there were no differences between the two groups in the teachers' giving the answer to the students.

Improvement resulting from sustaining feedback. When sustaining feedback was used, its immediate effect was noted by the observer as either improvement or no improvement. Improvement did not necessarily mean that the final desired answer was reached in the next interaction, but that the student had answered some question correctly or had otherwise improved his previous response.

When the absolute rates per minute of improved and not improved responses were examined, no significant differences were found between the two groups (37, 38; means = .08 and .13 per minute for improved responses, and .04 and .05 for unimproved responses).

As an index of the teacher's relative effectiveness using sustaining feedback, the proportion of all sustaining feedback which led to an improved response was computed. There were no significant differences between the two groups on this measure for total interactions (934; means = .68 and .73), and turn interactions (935; means = .73 and .77). However, there was a significant difference for nonturn interactions (936; $p = .04$, means = .63 and .70). These results suggest that the treatment teachers were slightly more effective with sustaining feedback, at least in nonturn interactions, although all teachers were generally successful in improving responses.

Each type of answer was examined for the proportion of *all* such answers which were improved, and the proportion which were sustained and improved. For incorrect answers, the proportion improved was significantly different for the two groups (937; $p < .01$, means = .27 and .39). However, there were no differences between the two groups for the relative effectiveness of sustaining feedback to incorrect answers (938; means = .68 and .72 of all *sustained* incorrect answers improved). Therefore, the treatment group's higher level of improved incorrect answers was probably the result of higher use of sustaining feedback, and not just better use of sustaining feedback when it did occur.

When *don't know* answers were analyzed, there was a significant difference between the groups for the proportion of all *don't know* answers which was improved (939; $p = .03$, means = .19 and .32). When examining only *don't know* answers which were given sustaining feedback, the proportion of these which were improved was also significantly different for the two groups (940; $p = .08$, means = .53 and .66). Not only did the treatment teachers tend to use more sustaining feedback in response to *don't know* answers, leading to more improvement overall, but they tended to be slightly more effective with such sustaining feedback when it was given.

When *no response* answers were examined, there was a difference for the proportion which was improved (941; $p < .01$, means = .21 and .40). However, there was no difference in the relative effectiveness of the treatment and control teachers in their use of sustaining feedback to *no response* answers (942; means = .68 and .74).

To summarize the variables describing improvement, the treatment teachers were sometimes slightly more effective in getting improved responses with their sustaining feedback, but this difference was not as notable as their greater use of such feedback.

Many times, teachers' use of sustaining feedback led to a sequence of interactions in which the teacher continued to give clues until the student got the answer right or until terminal feedback was eventually provided. In order to describe the length of such sequences, a variable was computed which measured the proportion of sustained interactions which received terminal feedback rather than further sustaining feedback. This represents interactions that were attempts to improve a child's answer but did not result in the final answer desired by the teacher, so that she had the choice of either continuing the sequence further or terminating. Therefore, the higher the proportion, the more likely was the teacher to terminate before success.

For total interactions, there was a significant difference between the groups for the proportion of sustained interactions which were terminated by the teacher (662; $p = .03$, means = .45 and .35). However, when broken into turn interactions and nonturn interactions, there were no significant differences (663, 664; means for turn interactions were .49 and .37, and means for nonturn interactions were .45 and .37). These data suggest that overall, the treatment teachers were likely to continue a sustaining sequence longer than the control teachers. That is, they were less likely to use terminal feedback within such a sequence and more likely to give further sustaining feedback, although they frequently did terminate before success.

Each type of sustaining feedback was analyzed to see how often it led to an improved response, and the two groups of teachers were compared to see if one was more effective with a particular type of feedback than another. There was no significant difference between the two groups for the proportion of repeating the question feedback which led to improvement for any type of interaction: total (944; means = .68 and .69), turn (945; means = .74 and .78), or nonturn interactions (946; means = .63 and .65). These means indicate that repeating the question usually was successful when used, especially in reading turns.

When examining the proportion of clue feedback which led to an improved response, no differences were found between the two groups for total interactions (947; means = .67 and .72), or turn interactions (948; means = .71 and .74). However, there was a significant difference between the two groups for nonturn interactions (949; p = .04, means = .62 and .71). When treatment teachers used clues in nonturn interactions, they were relatively more effective.

There were no differences between the two groups for the proportion of giving by clue feedback which led to improvement (950, 951; for turn interactions, means = .79 and .82, and for nonturn interactions they were .81 and .86). These results suggest that this type of sustaining feedback was easiest, as expected, although it did not absolutely guarantee a correct answer.

Summary of results for principles related to feedback to answers that are not correct (Principles 17,18, & 19). This group of principles as a whole had the strongest implementation of any in the instructional model. There was a basic pattern observed in all three types of answers (incorrect, no response and don't know) and both major types of questions (reading and nonreading). As expected, treatment teachers used more sustaining feedback, especially clues, and used less feedback which involved asking another student for the answer. There were fewer instances of other students calling out feedback in the treatment classrooms. The two groups were similar in their rate of giving the answer, especially during reading turn interactions.

In general, the differences between the two groups were strongest in interactions occurring outside of reading turns, during question and answer sequences. It is probably during such interactions that the teacher can exert more choice about what kind of feedback to use, since the pace is different. In oral reading turns, correcting mistakes as quickly as possible may be necessary to prevent interruptions of sentence or story meaning.

The treatment teachers were also slightly more effective with sustaining feedback than were the control teachers, when effectiveness was defined as yielding some improvement in the next answer. It is possible that they were concentrating more on the purpose of sustaining feedback, since the treatment materials had discussed why it should be beneficial. Perhaps this is a case of a self-fulfilling prophecy on the part of the treatment teachers: they believed that sustaining feedback should help, and therefore they were working harder at using it, and, in the process, made it more useful to the students. The control teachers, on the other hand, may have had no particular expectations about the efficacy of the technique, and therefore may not have tried so hard to be effective with it. This difference between the groups is interesting because the purpose of the treatment was to increase the frequency of sustaining feedback. Nothing was said about ways of using it more or less effectively.

Responding to Correct Answers (Principle 20)

Principle 20 suggested to the teachers that they should respond to correct answers by giving some kind of acknowledging feedba[illegible] and by making sure that all of the students heard and understood the answer. It was suggested to the teachers that they should, when necessary, repeat the answer or have it repeated, although this should not be overdone. Implementation of this principle was measured by noting the proportion of time that correct answers were followed by *no feedback*, and the proportion of time that the teacher *emphasized* the answer by repeating it or having it repeated. It was expected that treatment teachers would have a lower rate of *no feedback* and a higher rate of *emphasis* feedback.

There were no significant differences between the two groups in the absolute rate of *no feedback* (24; means = .06 and .10) or *emphasis* (25; means = .26 and .23).

The proportion of correct answers receiving *no feedback* was significantly different for total interactions (778; p = .10, means = .04 and .10) and for nonturn interactions (779; p = .09, means = .05 and .11). These results are not in the direction expected, because the treatment teachers were giving feedback to correct answers *less* often than the control teachers were. Indeed, the treatment group's means indicate that about one of every ten correct answers was not even acknowledged by the teacher.

The proportion of correct answers which received *emphasis* was significantly different for total interactions (780; p = .10, means = .40 and .31), and for nonturn interactions (782; p = .07, means = .43 and .33). However, there was no significant difference between the two groups in the proportion of correct answers receiving *emphasis* in turn interactions (781; means = .09 and .04). Again, this result was unexpected. Control teachers were emphasizing correct answers more often than treatment teachers.

Although the treatment did not recommend the use of emphasis for every answer, it was hypothesized that there would be a greater frequency of this for the treatment group, because the treatment had emphasized making sure the students knew what the correct answer was. Perhaps the treatment did make the teachers more aware of problems involved with repeating answers. If so, it might have served to decrease their use of this technique.

The observers also noted the teachers' responses to reading turns when they were completed correctly. There was no difference between the two groups in the use of no feedback in this situation (853; means = .22 and .24) or in the use of emphasis feedback (854; means = .12 and .08). It is interesting to note that, for both groups, over a fifth of all correct reading turns were not acknowledged at the end by the teacher.

In order to determine if feedback to correct answers was different according to type of question, the proportion of correct reading questions which received no feedback was computed. There were no significant differences for this variable for either total interactions (861) or nonturn interactions (862), although the means are ordered as they were for other variables, with treatment teachers more likely to omit feedback.

The proportion of correct answers to reading questions which received emphasis was not different for the two groups for total interactions (863; means = .37 and .27), or turn interactions (864; means = .09 and .04). However, there was a significant difference in nonturn interactions (865; p = .06, means = .41 and .31). Again, the control teachers emphasized correct answers more.

However, when correct answers to nonreading questions were examined, there were no differences between the two groups for measures of no feedback,

100

(908, 909) or for use of *emphasis* feedback (910, 911).

Summary of results for feedback to correct answers (Principle 20). Principle 20 was not implemented as expected. In fact, the results were opposite to those predicted.

Treatment teachers were more likely than control teachers to fail to give feedback in total and nonturn interactions. However, the difference was not highly significant, and it did not hold up when examined separately for different types of questions. One possible explanation of these unexpected results is that feedback to correct answers is not as important as originally believed, and that teachers were using it only when they felt it was necessary. Perhaps there was something about the treatment classes that made feedback to correct answers less necessary than in the control classes. However, there is nothing apparent in the treatment that could account for this.

The control teachers used *emphasis* more often in total and nonturn interactions, although the results were significant only at the .10 level. When broken down by question type, this pattern was maintained only for nonturn interactions with reading questions. The treatment did not specify that *emphasis* should follow every correct answer, but only that the teacher should make sure that everyone heard. In fact, it mentioned that there could be problems with too much emphasis by repetition. Perhaps this made the treatment teachers more aware of potential problems, and therefore lessened their tendency to emphasize answers.

Praise and Criticism (Principles 21 & 22)

Principles 21 and 22 made suggestions about the use of praise and criticism. It was expected that the treatment teachers would use less praise, and that they would be more specific with their praise and criticism.

Praise. The use of praise was noted in the low inference part of the coding system as a type of feedback to response opportunities, and as a type of teacher response to student behaviors.

When absolute rates per minute were examined, there were no significant differences for the rates of response opportunity praise (27; means = .13 and .08) or behavioral praise (425; means = .00 for both groups).

However, there was a significant difference between the two groups for the proportion of all contacts (response opportunities plus behavioral contacts) that involved praise (1033; p = .02, means = .12 and .07). When praise was examined just for response opportunities, there was a significant difference between the two groups for total interactions (730; p = .02, means = .14 and .07), and for nonturn interactions (732; p < .01, means = .16 and .07). However, there were no differences in the use of praise in turn interactions (731; means = .02 and .01). These results support the hypothesized treatment effect for lower incidence of praise in the treatment group, at least for nonturn interactions, where most praise was given. There was no significant difference between the two groups in the proportion of behavioral contacts that included praise (1014; means = .03 and .04).

When only correct answers were examined, there were significant differences between the two groups for the proportion of correct answers that were praised in total interactions (785; p = .01, means = .14 and .08), turn interactions (786; p = .09, means = .10 and .05), and nonturn interactions (787; p = .01, means = .15 and .08). Therefore, when only correct answers were considered, the treatment effect for the frequency of praise was stronger, with control teachers praising about twice as often as treatment teachers.

The proportion of reading turns that were completely correct and given praise was not significantly different for the two groups, although there was a trend in the same direction (855; means = .22 and .12).

When broken down by type of question, the proportion of correct answers to *reading* questions that received praise was significantly different for total interactions (868; $p < .01$, means = .20 and .09), for turn interactions (869; $p = .09$, means = .10 and .05), and nonturn interactions (870; $p < .01$, means = .21 and .09).

When *nonreading* questions in nonturn interactions were examined, the proportion of correct answers that received praise was not significantly different for the two groups (915; means = .07 and .05). Therefore, it can be concluded that the treatment effect was strongest for reading questions that led to correct answers.

The use of praise following incorrect answers was also examined. This occurred when teachers tried to praise some part of an answer, even though it was incorrect. The proportion of incorrect answers that received praise in total interactions was not significantly different for the two groups (799; means = .01 and .00), but there was a significant difference in nonturn interactions (800; $p = .09$, means = .01 and .00). Even though significant, the means do not indicate a large difference.

Reading turns (as a whole) that contained some mistakes were also praised some of the time, and the proportion of these which received praise did not differ significantly for the two groups, although there was a trend suggesting that control teachers gave more praise in this situation also

(859; means = .23 and .13).

Incorrect answers to reading questions were examined for the proportion followed by praise, and there were no significant differences between the two groups, for either total interactions or nonturn interactions (884, 885). In both cases, the means were .01 for the control group and .00 for the treatment group. It was not possible to analyze incorrect answers to nonreading questions that received praise, because this happened so seldom.

One of the more important points of principle 21 was that praise, when given, should be specific as to what was being praised. The proportion of all praise that was specific differed for the two groups (1034; $p = .04$, means = .03 and .07). When only academic praise was examined, the proportion that was specific also differed for the two groups (756; $p = .06$, means = .03 and .06). There was not a significant difference between the two groups in the proportion of behavioral praise that was specific (1029; means = .44 and .34). Perhaps behavioral praise is specific more often because it is typically given only when the teacher wants to point out a behavior with which he or she is pleased. Academic praise, on the other hand, tends to be a simple statement of "Good" following a response, and apparently it is meant to acknowledge and praise the correctness of the answer. It may be more difficult to specify, or perhaps it is less necessary to do so because the referent will be understood.

These data suggest that the treatment did have some effect in increasing the specificity of praise and decreasing the total amount of praise given by the treatment group. However, there was not high implementation of specificity by the treatment group, only greater implementation than the control group.

Criticism. When the two groups were examined for the absolute rate of academic criticism per unit of time, there was no significant difference, with both groups having a mean rate of .01 per minute (28). Likewise, there was no significant difference for the rate of behavioral criticism given, although the means were slightly higher (429; means = .03 and .02 per minute).

The proportion of all contacts that were critical in nature did not differ for the two groups, although the control group was slightly higher than the treatment group (1035; means = .08 and .05).

When response opportunities were examined, again there was no significant difference in the proportion of response opportunities receiving criticism for either total or nonturn interactions (733, 734). In both cases, for both groups, the mean was 1% of all response opportunities receiving criticism. Obviously, there was little academic criticism given in either group.

When only incorrect answers were examined, again there were no significant differences between the two groups for either total interactions (801; means = .01 and .02) or nonturn interactions (802; means = .02 and .03). When analyzed separately for reading questions and nonreading questions, again there were no significant differences between the groups for either total or nonturn interactions in the relative frequency of criticism to incorrect answers (886, 887, 921).

Principle 22 did not encourage criticism in the treatment teachers, so it is not surprising that there were no differences in the relative use of criticism. However, the principle did encourage a more specific use of criticism when it was offered. When all contacts involving criticism or correction of some kind were examined, there was no difference between the two groups in the proportion that was specific (1036; means = .06 and .07). This variable included not only academic criticism, but also correction delivered by the teacher for students' misbehavior. These corrections might or might not be considered critical, but they were definitely corrective in nature. One type, the behavioral management statement (a relatively mild correction), was examined separately for specificity. There were no significant differences between the two groups on the proportion of these that were specific, with means for both groups being .09 (1030). The proportion of total behavioral contacts that were specific demonstrated no group difference, with the means being .10 for each group (1019).

These data suggest that there was some natural implementation of the principle in the control group (6% of all critical and corrective comments were specific, and 10% of all behavioral corrections were specific), but the treatment did not increase the behavior of the treatment teachers above this base rate. Therefore, there was no evident treatment effect for Principle 22 and its suggestions about use of criticism.

Summary of results for praise and criticism (Principle 21 & 22). Principle 21, describing ways of using praise, was implemented to some extent. Treatment teachers used less praise, which is interpreted to mean a more moderate amount of praise. This was most apparent following correct answers to reading questions. There was some treatment effect for use of specificity with praise, in that the treatment teachers were more specific, but they did not use specificity very often, so this cannot be interpreted to mean strong implementation of the principle.

Principle 22 suggested that criticism should be very specific when delivered. There were no differences between the two groups for any measures of this. Control teachers were already being specific about 10% of the time, and the treatment teachers' use of specificity was no greater than theirs.

Many other variables were derived from the observation system which did not directly measure implementation. However, in order to determine if the two groups of teachers differed in other ways, other groups of variables are also reported.

Time Usage

There were no significant differences between the two groups for either the average total time for reading groups, which included transition time plus R.O. time (4035; means = 26.26 minutes for control and 26.85 minutes for treatment), or for the average response opportunity time available within that period (4029; means = 22.53 minutes and 24.02 minutes).

Several variables indicated the rate at which the teacher gave response opportunities of various kinds to the students. Some of these have already been reported as measuring implementation of Principle 7: total R.O.'s per minute, (601; $p = .07$, means = 1.86 and 2.20 per minute), and nonturn R.O.'s per minute (602; $p = .02$, means = 1.20 and 1.64 per minute). There were no significant differences between the two groups of teachers in the rate at which they offered reading turns (603; means = .40 and .36 reading turns per minute), or in the overall rate per minute of response opportunities during reading turns (604; means = .65 and .56).

There was a significant difference between the two groups in the proportion of all response opportunities that occurred during reading turns (605; $p = .05$, means = .36 and .26). However, within each reading turn, there were no differences between the average number of response opportunities per turn (606; means = 1.66 and 1.56). Therefore, the difference between the two groups for the proportion of total response opportunities that occurred during reading turns is probably due to the higher frequency of nonturn response opportunities per minute in the treatment group. This has already been discussed with regard to Principle 7.

Because the treatment group was using more sustaining feedback, and because each use of sustaining feedback led to another response opportunity, it might be argued that the higher frequency of response opportunities was due to a higher rate of sustaining feedback. In fact, the proportion of all response opportunities which were initial selections rather than sustained selections showed no differences between the two groups (645; means for each group = .66). The other 34% of response opportunities (which were not initial selections) were due either to sustaining feedback or to the teacher continuing with the student by asking a new question after giving an answer through terminal feedback.

To summarize data on the ways that time was spent in the reading groups, it can be seen that, on the average, teachers in the two groups spent about the same amount of total time teaching reading in their groups, but the treatment teachers managed to present many more response opportunities in nonturn situations. This means that they were asking more single questions, but they were not having students read aloud more times than the control teachers.

Observers noted the lesson contexts in which response opportunities occurred. Five contexts were used: a) slow-paced questioning and answering without use of a workbook, a worksheet, or basal reader; b) workbook or worksheet activities; c) fast paced drill; d) reading a story aloud from the basal for the first time; e) reading a story aloud from the basal that has been read previously in the group. There were no significant differences between the two groups for the average time spent in any context or the proportion of response opportunity times spent in any context (4030-4034, 4063-4067).

Whenever a teacher was in the process of teaching a lesson and had to leave the group for any reason, the observer noted this and timed her absence. It was thought that this measure might be reflective of teachers' overall management ability. There were no differences between the two groups for either the average time the teacher was out of the group when she had to leave, or the average number of times per observation that the teacher left the group (4036, 4037).

There were significant differences between the treatment and control teachers for the number of groups seen during a morning (4069; $p = .07$, means = 2.95 for control and 3.41 for treatment), and the average group size (4070; $p = .09$, means = 7.55 for control and 6.50 for treatment). These results indicate the treatment teachers were more likely to see fewer children in more groups than were the control teachers, on the average. Because the treatment said nothing about optimal size of groups, this difference cannot be directly related to it. In fact, group assignments were made before the treatment was introduced.

There was a significant difference between the two groups for the average number of activities that were given to the students to do either during a group lesson or at their seats after the lesson (5352; $p = .03$, means = 1.02 for control and 1.80 for treatment). The treatment teachers assigned slightly more activities to their students. (This measure does not take into account length or difficulty of activity, but only the number of different activities.)

These data on time usage suggest that the treatment teachers may have been more demanding of their students. Because groups were slightly smaller on the average in treatment classrooms, and because treatment teachers had higher rates of questioning, this may mean that their students were getting exposed to more opportunities for practice of skills in the group. This, combined with slightly more follow-up work, might mean that treatment students, on the average, were exposed to more content that demanded their direct attention by requiring them to answer questions or complete assignments.

Part of this could be related to the treatment (more practice by individuals in the group), but other parts cannot (size of group, number of assignments).

Summary of Time Usage Data. On the average, control and treatment teachers spent the same amount of total time teaching reading in their groups. They also had similar rates of use of each of the five lesson contexts which were measured. However, treatment teachers had higher rates of response opportunities per time, especially nonturn interactions, and they also had slightly more activities assigned to the students to follow up on the reading group lesson. Treatment teachers had smaller groups and more groups than control teachers. Some of these differences can be related to the treatment (more response opportunities given to individuals), but other group differences could not be attributed to the treatment (the size of the group, or the number of assignments given to the students).

Curriculum Used and Content Covered

Although the instructional model was considered to be "curriculum-free", information was gathered about the materials used by the teachers and how much was covered during the year in order to determine the relationship of this to achievement. It would be expected that there would be few differences between the treatment and control group on the choice of materials used, since schools were assigned randomly to treatment.

The variables described in this section were computed by reading group rather than by teacher, since the curriculum used with each group varied even within classrooms. Therefore, these analyses are based on 72 reading groups, 34 in control classrooms and 38 in treatment classrooms.

There were three basal series adopted by the school district for use in first-grade classes: Economy, Harcourt-Brace, and Houghton-Mifflin. Unexpectedly, there were significant differences between the treatment and control groups on the principal basal series used. The Economy series was used significantly more often in treatment reading groups than in control reading groups (6002; $\underline{p} < .01$, means = .24, .63). Therefore, more than half of the treatment reading groups used Economy, while only about a quarter of the control reading groups used this series. Although all of the series included phonic skills, these were emphasized more heavily in the Economy series.

Harcourt-Brace was used more frequently in the control reading groups (6003; $\underline{p} = .01$, means = .30, .08), as was Houghton-Mifflin (6004; $\underline{p} = .05$, means = .45, .24).

These differences in choice of basals may represent a school effect.

Since treatment and school were confounded, this effect is undesirable, but was unavoidable.

Other measures of content covered looked at the reading level completed by the end of the year and the number of basals covered during the year. These comparisons showed differences, although not highly significant, between the treatment and control classes. The treatment group finished at a slightly higher reading level (6001; $p = .10$, means = 2.50, 2.89, where a 3 on the scale represented the primer in the particular series). This suggests that these first-grade reading groups were not covering as much material on the average as might be expected, since the first reader (4 on this scale) is considered to be the target for the end of first grade. The number of basals completed during the year was also examined. This is similar to the variable just reported, except that the former did not take entering level into account. The "number of basals completed" reflected the number of books the students completed, regardless of where they started. There was also a weak significant difference here between the two groups, with the treatment reading groups completing slightly more than the control group (6006; $p = .06$, means = 4.5, 5.1). This result suggests that the treatment teachers were moving their reading groups through the basal at a slightly more rapid pace.

Other measures indicated whether or not the teacher used additional standard curriculum materials with the students, such as workbooks. There were no significant differences between the treatment and control group for systematic use of basal workbooks (6007; means = .91, .87, indicating that almost all reading groups did use the workbooks). However, there were differences favoring the treatment group for use of other supplementary materials: worksheets accompanying the basal readers (6008; p = .06, means = .06, .21) and spelling workbooks (6009; p = .03, means = .21, .45). There were no significant differences for the use of a handwriting workbook (6010; means = .21, .18). There were significant differences between the groups for the use of an English workbook, but these favored the control group (6011; p = .03, means = .12, .00).

There were no differences in the use of the DISTAR program, with only a few control reading groups using this (6012). There were significant differences for the use of SRA materials, with only a few treatment reading groups using these (6013; p = .03, means = .00, .13).

When the average size of the reading group was computed with groups as the unit of analysis, the difference between the treatment and control groups was more significant than that found when reading group size was computed as a class mean. Again, the control group showed larger reading groups on the average (6014; p = .01, means = 8.06, 6.40). When average class size over the entire year was examined, there were similar differences, with control classes being larger (6019; $p < .01$, means = 26.53, 22.79).

During the year, the observers kept track of changes in reading group membership, and computed various indices of the stability of the group. There were significant differences between treatment and control classes for the average number of changes during the year in reading groups (6015; p = .02, means = 7.00, 4.30). Therefore, treatment reading groups had fewer additions or changes. It is not known whether this was due to less mobility in and out of the school, or if it was due to control teachers' more frequently rearranging the reading group within the classroom. An overall rating of the stability of reading group membership was made, based on a ratio of changes during the year to the average group size. (A *1* indicated very stable membership and a *5* indicated very unstable group membership.) This rating of stability of reading group membership showed no significant differences between treatment and control classes, with both having an average rating in the middle of the scale (6016; means = 2.59, 2.37). This suggests that the greater number of moves in the control classes may have been at least partially due to their larger size. (That is, there were the same number of moves per pupil.)

Summary of group differences on curriculum and content covered. Although it was not expected, these analyses yielded several differences between the treatment and control classes. Treatment teachers used the Economy reading series (with heavier emphasis on phonetics) more often, and control teachers were more likely to use Houghton-Mifflin and then Harcourt-Brace with their reading groups. The treatment teachers covered slightly more basal material during the year, and were at a slightly higher reading level at the end of the year, although these results were not highly significant. There were also some indications that the treatment teachers used additional materials more than the control teachers, such as commercial worksheets, a spelling workbook, and SRA materials. The control teachers used the English workbook more, although not many of them used it. Control classes were larger on the average and had larger reading groups than treatment classes, and there were also more changes among reading group membership in the control classes, although this was probably due to their greater size, since a rating scale which controlled for group size indicated no differences between the treatment and control groups on relative stability.

Other Categories of Academic Teacher-student Interaction

When describing response opportunities, the observer noted for each interaction the type of selection, the type of question, the response of the child, and type of feedback given. Most selection and feedback variables have already been discussed, because they directly measured implementation of some of the principles. However, variables describing types of questions and levels of answer were also computed.

The observer described each interaction in terms of one of nine types of questions. As discussed earlier, these could be classified as either reading questions, which required the student to decode words or provide information about sound and letters, and nonreading questions, which asked the student to provide information about something that had been read, or an idea that was being discussed. Reading questions were:

1. repetition - the teacher asked the student simply to repeat something she said.
2. reading choice - the student had to decide between some definite alternatives that had to do with decoding words or sounds.
3. word recognition - the student was to look at a word or letter and say its name.
4. word attack - the student had to answer a question about a part of a word or the sound made by a letter or letters.

Nonreading questions were:

5. personal - the student was asked to provide information about an opinion or personal experience.

6. choice - the student had to choose between some definite alternatives in order to answer a question that did not require him to decode a word.
7. product - the student was required to give a factual answer to a question that was not a comprehension question.
8. comprehension - the student was asked to give a factual response based on some material that had just been read.
9. interpretation - the student was asked a "why" or "how" question that did not involve decoding a word.

When the absolute rate per minute of types of questions was examined, there were no significant differences between the two groups except for the absolute rate of repetition questions and comprehension questions (11-19). Although the rate of repetition questions produced a significant difference (11; $\underline{p} = .10$), the means were so low that it was impossible to tell the direction (.00 for each group). The treatment group had a higher absolute rate of occurrence of comprehension questions than the control group (18; $\underline{p} = .10$, means = .08 and .15 per minute). However, the proportions of response opportunities which were of each type of question showed no significant differences between the two groups. This included variables that measured the proportion of response opportunities that were reading questions and nonreading questions (666-689).

Because one important part of the treatment encouraged the teachers to use sustaining feedback by asking simplifying questions, the types of questions used for sustaining feedback were compared. However, there were no significant differences in the types of questions used to correct errors (653-659).

These results indicate that the treatment had very little effect on the general types of questions asked by the two groups.

The treatment did not make any direct suggestions about level of difficulty or error rate, except to suggest that simpler, easy-to-answer questions should follow initial mistakes. However, the overall error rates were significantly different for the two groups.

When the absolute rate per minute of correct, incorrect, don't know, and no response answers were examined (20-23), there were no significant differences except for the occurrence of don't know responses (22; p = .07, means = .01 and .02 per minute). Although this was a statistically significant difference, the means do not indicate a meaningful difference between the two groups in rate of occurrence.

The proportion of all response opportunities for total interactions which involved correct answers was significantly different for the two groups (690; p = .04, means = .66 and .73).

Likewise, the proportion of total response opportunities that were incorrect answers was significantly different for the two groups (693; p = .04, means = .27 and .21).

There were no differences between the two groups for the proportion of total response opportunities that included don't know answers (696; means = .01 and .02).

There was a significant difference between the two groups for the proportion of response opportunities which included no response answers (698; p = .06, means = .14 and .10).

However, when all types of answers were examined separately for non-turn interactions, there were no significant differences (691, 694, 697, 699).

There were no significant differences between the groups in the number of mistakes made during oral reading turns that led to interactions with the teacher, although there was a near-significant trend for the control teachers to have slightly more errors (701; means = .79 and .60 mistakes per reading turn).

There was also a near-significant trend for differences in the proportions of reading turns that were completely correct (702; means = .60 and .68).

These two trends suggest that the treatment teachers' students made fewer mistakes in oral reading turns. This probably contributed to the significant differences reported for the proportion of total R.O.'s which included correct or incorrect answers (690, 693), since there were no differences between the groups for the level of answer in nonturn interactions.

Variables that represented combinations of types of questions and types of answers were examined to see if teachers in the two groups were differentially successful with different types of questions.

The proportion of total reading questions which led to correct answers was significantly different (952: $p = .05$, means = .62 and .70). However, when broken into turn and nonturn reading questions, there were no differences between the two groups (953, 954). There also were no significant differences for the proportion of nonreading questions that lead to correct answers (956). The only other significant finding for this type of variable was for the proportion of total word recognition questions that were correct (708; $p = .02$, means = .50 and .62). When broken into word recognition questions occurring during turn and nonturn interactions, there were no significant differences (709, 710). Differences for other types of questions also were not significant (706-722).

To determine whether the treatment teachers' higher rate of correct answers was a reflection of their higher use of sustaining feedback, the proportion of initial selections which led to correct answers was examined. These variables represent an error rate for questions asked to a student at the beginning of a sequence of interactions, and therefore would not include any correct answers resulting from sustaining feedback with simpler questions.

For each of the five types of initial selection (ordered, preselect, nonvolunteer, volunteer, and call out), the treatment teachers had a slightly higher rate of correct answers (640-644). However, this was significant only for nonvolunteer selections (642; $p = .05$, means = .66 and .72 of such selections leading to correct answers).

These data suggest that the treatment teachers, on the average, had a relatively higher rate of correct answers, and that this cannot be attributed to their greater use of sustaining feedback. The differences between the groups were not extreme, however. Both groups had more correct than incorrect answers. For example, when all interactions were considered, the control teachers averaged 66% correct answers, while the treatment teachers averaged 73%.

Two other types of feedback were observed and coded that did not relate directly to any of the treatment principles but might have had indirect relationships. These two were process feedback and new question feedback. In process feedback, the teacher gave a brief discussion of the process used to arrive at a correct answer, or an explanation as to why an answer was wrong. A new question was the type of feedback noted when the student was allowed to continue a sequence of interactions with the teacher, but about a brand new question. This meant that the previous question had been answered, either by the student or by the teacher. It is to be distinguished from sustaining feedback following errors, in which the purpose is to address the original question.

When process feedback was examined, there were no significant differences between the two groups for the absolute rate of occurrence (26) the proportion of correct answers receiving process feedback (728, 729, 783, 784, 866, 867, 912, 913) or the proportion of incorrect answers receiving process feedback (797, 798, 882, 883, 918, 919).

There also were no significant differences between the two groups for the absolute rate of new question feedback used (33) or the proportion of interactions which included new questions (743, 745, 788-790, 856, 871-873, 916, 917, 811-813, 860, 896-898, 928, 929, 841-843). Neither of these types of feedback was discussed in the instructional model, so it was not surprising that there were no significant differences between the two groups.

Other variables were computed to describe the use of a new question following terminal feedback, because it was felt that teachers who might otherwise use sustaining feedback to follow errors, but who occasionally had to use terminal feedback, would be more likely to extend a child's response opportunity by asking a related new question. However, there were no differences between the two groups for use of a new question with terminal feedback (957-959).

Summary of results of group differences for other response opportunity categories. Other categories describing response opportunities were examined for differences between the treatment and control groups in types of questions asked, rate of correct answers, and use of other types of feedback. Since none of these were emphasized in the instructional model, no differences were expected. Several types of questions were examined for absolute and relative frequency, but there were no significant differences between the two groups. However, there were differences in relative error rate. Although all teachers usually asked questions that led to correct answers, the treatment teachers had a higher rate of correct answers, and fewer failures to respond. There were no differences between the two groups in their use of two other types of feedback: process feedback (an explanation of how to find the answer) or new question feedback (a new question is asked of the same child).

Behavior Contacts

The last set of variables to be presented describe the types of behavioral corrections given by teachers to students who misbehaved while a group was being conducted. There was no significant difference between the two groups for the proportion of all contacts that were behavioral contacts (1031; means = .18 and .13).

The types of misbehaviors that were corrected were recorded for each behavioral interaction. There were no significant differences for the absolute rate of occurrence of any type of misbehavior (413-424), and only one significant difference in relative rates (1002-1009), dealing with the proportion of behavior contacts that were for possession of contraband (1009). These were corrections that resulted when students had something in the reading group with them that was not allowed, such as a rubber band or toy (p = .08, means = .02 and .01). Although this is statistically significant, the means indicate that the actual difference was very small.

There also were no differences between the two groups for the proportion of behavioral contacts that were noninteractive and nondisruptive (1012; means = .73 and .67) or interactive, potentially disruptive, and therefore more serious (1011; means = .27 and .32). These means indicate that both groups of teachers used most of their corrections for relatively minor misbehavior, although both also had some problems with potentially disruptive misbehaviors.

Each behavioral contact was described in terms of the teacher reaction to the misbehavior. There were no significant group differences for any type of teacher reaction (1015-1018, 1025-1028).

Behavioral contacts were classified as being in-group (the teacher initiated some correction with a student in the reading group at that time) or out-of-group (either a student interrupted the teacher or the teacher interrupted the group lesson to correct someone at his seat). There were no differences between the two groups of teachers for the proportion of all corrections which were in-group rather than out-of-group (1040; means = .50 and .47). There was a significant difference in the proportion of all contacts (response opportunities plus behavioral contacts) that were in-group corrections (1077; p = .10, means = .09 and .06).

There were no differences in the proportion of all contacts which were out-of- group contacts (1078; means = .09 and .07). There was a significant difference for the proportion of all teacher-initiated contacts (this included response opportunities, in-group corrective contacts, and out-of-group contacts initiated by the teacher) that were teacher-initiated out-of-group contacts (1079; p = .07, means = .06 and .04).

These data suggest that the control teachers may have had slightly more behavior problems while teaching their reading groups than the treatment teachers, although the differences are small. It is tempting to attribute this to the differences in the teachers' seating themselves so as to monitor the group and the class more carefully, and the emphasis in the treatment as a whole on teacher control of the lesson. However, this interpretation cannot be advanced until there are further analyses relating classroom processes to one another.

Variables were created to indicate the length of most of the out-of-group contacts. When only child-initiated out-of-group contacts were considered, there were no significant differences between the groups for the proportion that were brief in duration rather than long (1041). The same is true for only teacher-initiated out-of-group contacts (1042). There was a significant difference between the two groups in the proportion out-of-group contacts that were child-initiated rather than teacher-initiated (1043; p = .02, means = .32 and .50). These data suggest that control teachers interrupted their own reading groups more than outside students did.

There were no significant differences between the two groups in the ways that teachers dealt with child-initiated out-of-group interruptions, or the ways they contacted students who were out of the group (1045-1076).

Overall, the data for behavioral corrections suggest that the control teachers may have spent slightly more time and attention on control problems, although this conclusion is based primarily on results significant only at the .10 level.

Summary of group differences on behavior contacts. The control teachers had slightly more behavior contacts than did the treatment teachers. They were more likely than the treatment teachers to interrupt their instruction to deal with students out of the group who were doing independent work. However, there were no differences between the two groups for the types of misbehaviors that occurred or the teacher reaction to them.

Discussion

Many of the behaviors suggested by the instructional model were implemented by the treatment teachers at a significantly higher rate than that of the control teachers. Some of the principles were not implemented at all by the treatment group, and some of them were already being used by the control group to the same extent as the treatment group.

The principles that showed the strongest treatment effect on teacher behaviors dealt with discrete, easily described behaviors that were probably already in the teachers' general repertoires of techniques (e.g., maximizing use of ordered selection and minimizing use of volunteers and call outs; maximizing appropriate use of sustaining feedback and minimizing use of asking another student for the answer). Although teachers might not have used these techniques extensively before the treatment, the description of the required behaviors was apparently understood. The teachers could easily analyze their own behaviors in these terms and monitor their use. They were also apparently convinced by the rationales for their inclusion in the treatment. These were behaviors suggested by earlier research in similar classroom settings, and so it is reasonable to assume that the teachers recognize their potential value.

On the other hand, behaviors that showed no treatment effect and were not highly implemented by either group were less specific and possibly novel to the teachers (e.g., use of a model, breaking up the group, use of a signal before lessons). The description and rationale for them was not sufficient to cause implementation. It cannot be determined from the data here if the failure to implement was due to lack of specificity, novelty, lack of sufficient rationale, or inappropriateness for the setting.

Other behaviors were used to some extent by the control group, and the treatment did not increase the level of use by the treatment teachers. Examples of such behaviors were using a signal to start the transition and sitting in position to be able to monitor the entire class. Lack of significant differences for these variables might indicate that most teachers recognize the efficacy of the principles and are already implementing them regularly.

Even when a treatment effect was found, it was not an "all or nothing" phenomenon. That is, the treatment teachers never used a suggested technique all of the time, and the control teachers always used it some of the time. It is important to recognize this in building realistic treatment programs which acknowledge that teachers must use their own judgment about when the situation warrants a particular behavior. For example, sometimes a strong treatment effect was found only for interactions occurring outside of reading turns, where the pace is slower and the teacher's options for questions and feedback are greater. There are probably many other important types of contextual influences to consider in studying such classroom processes, especially when trying to bring about change in those processes. Some of these contextual influences are discussed in Anderson, Morgan, Evertson, and Brophy (Note 5).

In summary, the treatment was generally successful in influencing change in the treatment teachers' behaviors, but the results demonstrated that not all components of the treatment were equally successful. An analysis of the different results for these parts suggests that future treatment models should consider the specificity and familiarity of the behaviors, their appropriateness for particular contexts, and the role of teacher judgment in implementing them.

These analyses also revealed that the treatment teachers differed from the control teachers in ways that could not be attributed directly to the treatment. In trying to account for the treatment effect on overall achievement, these differences must be considered. The possibility exists that the treatment teachers were different from the control teachers before the study began, despite random assignment of schools to groups, and there is also a possibility of a Hawthorne or expectancy effect, which influenced the treatment teachers to work harder and do a better job simply because they knew they were an experimental group and were expected to do better. These possibilities are discussed further in Chapter 5, where the group differences are compared to the regression data to see how closely they match, in order to determine whether the achievement differences were due to the treatment.

Chapter 4: Results of Regression Analyses

This chapter presents relationships between the classroom process variables and student achievement. In all analyses (except for one set, which is discussed in the text), the class mean scores were used to represent the average readiness level of the students at the beginning of the year, the average achievement of the students on two tests given at the end of the year, and the average score for that class or teacher on each process variable. Therefore, the N for all analyses (unless indicated otherwise) was 20.

The analyses used to determine the strength and direction of process-product relationships were a series of comparisons of linear models. This approach to regression analysis is described in Ward and Jennings (1973), and the computer programs used to create the particular models used for this study were developed by Veldman and Linsley (Note 6). The models used and the hypotheses tested in each comparison are described below, and an explanation is given of the tables, which are reduced from the computer output.

Data Analysis

Two sets of linear regression equations were compared for each of the potentially predictive teacher or classroom behavior variables. One set provides the degree of simple relationship to gain and also the degree of interaction with initial ability. The second set of equations identifies the extent and nature of any second-degree curvilinear (quadratic) relationships between the variables. These analyses are included in the tables whenever there is an interpretable curvilinear effect.

Linear Relationships

The three regression equations used in this set are shown below. As indicated, each produces a squared multiple correlation coefficient, and comparisons of these yield two F-ratios and associated chance probability values.

$$\text{Ach} = \text{TRD} + \text{CB} + (\text{CB})\ (\text{TRD}) + E_1 \qquad R_1^2$$

$$\text{Ach} = \text{TRD} + \text{CB} + E_2 \qquad R_2^2$$

$$\text{Ach} = \text{TRD} + E_3 \qquad R_3^2$$

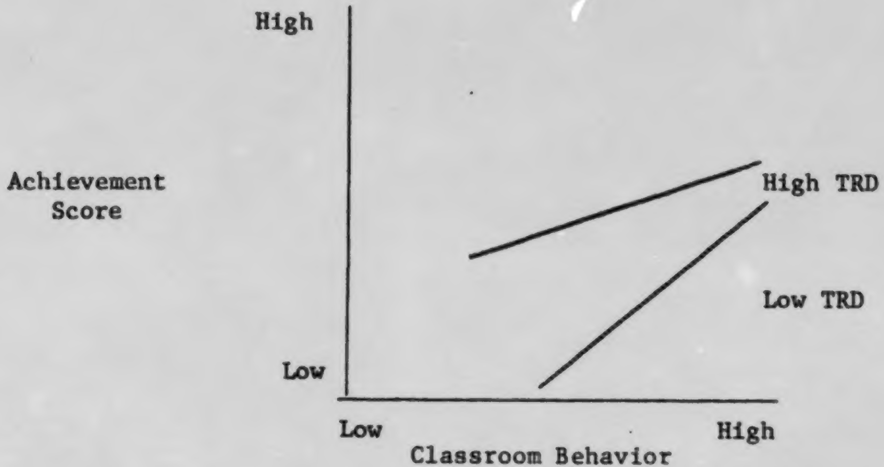
$$F_1 = \frac{(R_1^2 - R_2^2)\ (N - 4)}{(1 - R_1^2)} \qquad df = 1,\ (N - 4)$$

$$F_2 = \frac{(R_2^2 - R_3^2)\ (N - 3)}{(1 - R_2^2)} \qquad df = 1,\ (N - 3)$$

In these equations "Ach" is the Total Reading or Word Analysis achievement score from the MAT given at the end of the school year, "TRD" is the Metropolitan readiness measure, "CB" is the particular classroom behavior variable being assessed, and "E" represents errors of prediction. Each equation is solved for a set of weights which minimize the E values, thus maximizing R^2, which is an index of the amount of criterion variance associated with the predictor variables in the equation.

The first equation R^2 must equal or exceed that of the second, which must in turn equal or exceed that of the third, because each equation contains successively less information. The product variable in the first equation represents the interaction of initial ability and teacher or classroom behavior, and the first F-test therefore assesses whether the relationship is the same at all levels of entering readiness. The second model comparison forces the relationship to be common at all ability levels, and then asks whether the relationship is significantly greater than zero. Because the readiness score appears in all equations, it is said to be "statistically controlled." For instance, the second comparison asks whether achievement is predictable from the teacher behavior beyond what is predictable from the readiness score. In other language, we are asking whether classes that are exposed to different levels of the teacher or classroom behavior, but which are identical in initial readiness differ in their expected (predicted) achievement at the end of the year.

In the event that the interaction is found to be statistically significant ($\underline{p} \leq .10$), expected values for the achievement test are calculated for particular combinations of readiness level and classroom behavior in order to explicate the nature of the interaction.. Four combinations are sufficient: a) low TRD with low CB; b) low TRD with high CB; c) high TRD with low CB; d) high TRD with high CB where "high" and "low" are plus and minus 1 sigma from the mean of the variable concerned. To facilitate comparisons across classroom behavior variables, these values are scaled as $\underline{z}$ scores (mean = 0, SD = 1). In the example below, we see that the behavior is positively related to gain, but that it is more strongly related among classes initially low in readiness than among those initially high.



The second test of the series, which forces the implicit regression lines to be parallel, may or may not be significant, independent of the interaction effect. If both are significant, we can make a general statement about the classroom behavior's effect, but with a qualification recognizing its interaction with initial readiness.

In the event that only the second test is significant, we can determine the direction of the effect of the classroom behavior simply by examining the sign of the CB beta weight in the second equation.

Curvilinear Relationships

The previous set of models is sensitive only to the linear aspects of the relationship between classroom behavior and gain. To determine whether regression lines which are allowed to curve will better fit the actual data points, another set of regression models was employed.

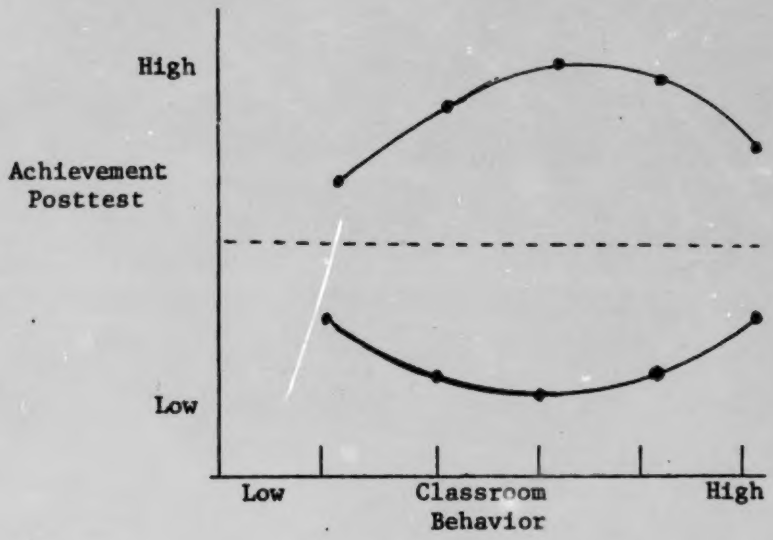
$$\text{Ach} = \text{TRD} + \text{CB} + (\text{CAT})\ (\text{TRD}) + (\text{CB})^2 + (\text{TRD})\ (\text{CB})^2 + E_1 \qquad R_1^2$$

$$\text{Ach} = \text{TRD} + \text{CB} + (\text{CAT})\ (\text{TRD}) + E_2 \qquad R_2^2$$

$$F = \frac{(R_1^2 - R_2^2)\ (N - 6)}{2(1 - R_1^2)} \qquad df = 2,\ (N - 6)$$

The second of these equations is, of course, the first of the previous set. By adding the last two terms - squared CB scores and their products with the readiness scores - we permit the lines not only to bend once, but to bend differently at different levels of the readiness test.

If the F test is significant, we conclude that allowing the regression lines to bend does indeed afford a better fit to the data. To obtain a graphic reflection of such an effect, five expected values are computed for the low readiness level and five also for the high readiness level. Classroom behavior values for the mean, plus and minus 1/2 sigma, and plus and minus 1 sigma are plugged into the equation with a high or low readiness score. The resulting set of ten values can be used to produce a plot such as this example:



In this example, the suggestion would be that in the midrange of the classroom behavior, it has a depressing effect on the performance of low readiness classes and an enhancing effect on those of high readiness classes, but when the classroom behavior is relatively high or low, readiness is not relevant. There is also the suggestion here, reflected by the dotted line, that for average ability classes, the classroom behavior is not associated with achievement at all.

Each variable describing classroom processes was analyzed in the manner shown below for two achievement scores: Total Reading (which was the sum of the Word Knowledge and Reading subtests) and the Word Analysis subtest. The tables are reproduced as they come from the computer printout (Veldman and Linsley, Note 6). The following output will be used as an example to aid the

Figure 1: Example of Tables Describing the Results of Regression Analyses

Variable label (Classroom Behavior - CB)

Criterion of interest (TLR = Total Reading scores, WDA = Word Analysis score).

Difference in R^2 between models 1 & 2 (test for interaction).

Difference in R^2 between models 2 & 3.

Beta weight represents change in criterion (z scaled) per one sigma increase in classroom behavior. Sign indicates direction of change. If interaction is significant, best prediction is obtained from model 1.

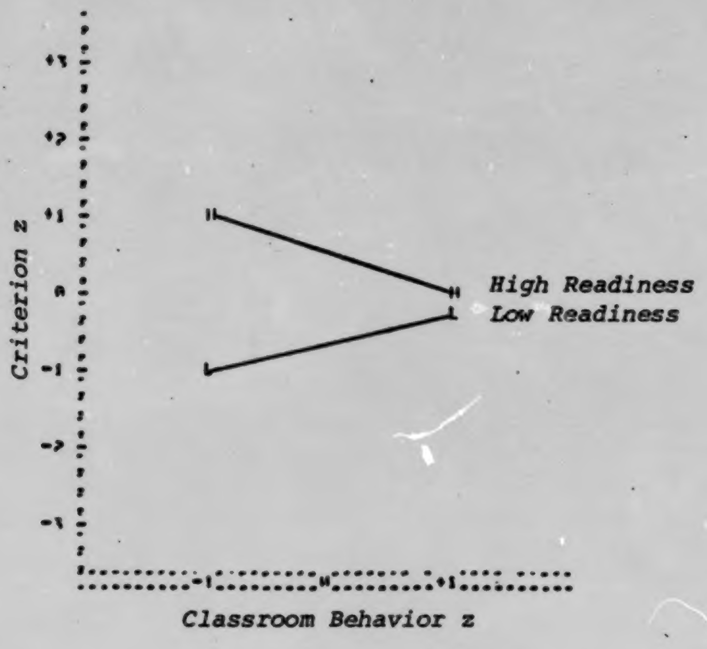
Range in raw score points of the classroom behavior at + or - 1 sigma.

```
NUMBER OF CALL OUTS PER MINUTE

        CB VARIABLE 27       ID =  8052

        MODEL 1. TLR = TRD + CB + TRD*CB    R2 =   .5702
           DIFFERENCE =   .1186    P =   .0095   **
        MODEL 2. TLR = TRD + CB               R2 =   .4517
           DIFFERENCE =   .0037    P =   .7363
        MODEL 3. TLR = TRD                    R2 =   .4479
        BETA FOR CB IN MODEL 2 =  -.0625
        HI CB =      .3392   LO CB =     .0568
        EXPECTED VALUES FROM MODEL 1. (Z-SCALED)
           LO TRD, LO CB =    -.8931
           LO TRD, HI CB =    -.2588
           HI TRD, LO CB =     .8926
           HI TRD, HI CB =    -.0621
```

Variable ID # used in the text.

p < .10 will result in the plotted interaction shown below.

p < .10 indicates a significant linear relationship between the classroom behavior and criterion.



reader in understanding the data tables.

In the text, the results are accompanied by some statistical information. When results are significant, the *p* level is given and the range of *plus and minus 1 sigma* is reported for the classroom behavior variable. Since two series of tests were run for each variable (one for each achievement score), both are reported at the same time to determine the strength of relationship. Since the *N* was low (20 classes for most analyses), and since patterns in the results were considered more important than the absolute probability level of any single test, a value of $p \leq .10$ was selected for considering the results to be significant. Since two criteria were available, the following rule was used to determine inclusion of results as significant. Results are reported as significant if the classroom behavior variable was related to *one* of the test scores with an associated *p* level of .05 or less, *or* if it was related to *both* test scores at a level of .10 or less. Therefore, results in which the variable was weakly related to one criteria (i.e., $.05 < p \leq .10$), but not related to the other ($p > .10$) were not reported as significant.

Variable numbers are given for all results so that the tables may be referenced (see below for table numbers including variables of interest). When results are significant, information is included in parentheses to give the variable number, the *p* levels for the two tests, with Total Reading first, then Word Analysis, and the range of the classroom behavior that fell within plus and minus 1 sigma from the mean. For example:

(4035; $p < .01$, $p = .05$; range = 21.24 to 31.84)

would indicate that variable 4035 (which was the average time that reading groups met) was significantly related to the Total Reading score and the Word Analysis score, and that most groups met from 21 to 32 minutes on the average. The discussion of this result in the text defines this relationship as being positive. The beta weight for the slope may be found in the tables.

Since they were analyzed separately, the results for the Total Reading scores and Word Analysis scores appear in different tables. The tables are also divided according to the type of data describing classroom processes in the same way that was done for the group comparison data. The following is a guide to finding variables in Tables 10 through 18:

Variable	Describing	Test Score Used as Criterion	Located in Table
4026-5369	Reading group measures	Total Reading	13
1-431	Rate variables	Total Reading	14
601-1079	Proportion variables	Total Reading	15
4026-5369	Reading group measures	Word Analysis	16
1-431	Rate variables	Word Analysis	17
601-1079	Proportion variables	Word Analysis	18

Throughout the text, reference is made to "low ability" or "low readiness" classes, or to "high ability" or "high readiness" classes. These refer to classes that were one sigma below or above the mean respectively on class mean readiness, as measured by the Metropolitan Readiness test. Whenever interactions were detected such that the relationship with achievement depended on both the level of the classroom behavior and the entering readiness score, the computer program plotted the relationships for such classes to illustrate the different slopes. For purposes of brevity, the text will refer to "low" and "high" classes, but the reader should understand that these references are to the statistical extremes, rather than to any absolute level of ability or readiness.

Results

Getting and Maintaining Students' Attention (Principles 1 & 2)

Use of an attention getter and its effectiveness. There were no significant relationships for the proportion of times that an attention getter was used to begin a transition (5312) or to begin a lesson (5321).

The variables describing the kinds of attention getters used in transitions and at the beginning of the lesson also showed no significant relationships for either test score (5313, 5315, 5316, 5318, 5319).

Although there were no significant findings for the use of signals to begin transitions, there were important relationships with achievement for the smoothness of transitions. There was a significant negative relationship for the proportion of total observed time that was spent in transition (4062: p's = .03, < .01, range = .07 to .15); therefore, the higher the proportion of total time spent in transition, the lower the achievement. It is likely that this variable reflects the teachers' overall management abilities: the more effective teachers had smoother and shorter transitions.

Other variables describing time spent in transition were the average time taken for each of three components of the transition: the time to the group for students, the time to the group for the teacher, and the time to the lesson once the group was together. For the first two variables, there were no significant relationships (4026, 4027). However, the average time taken for the lesson to begin once the students and teacher were in the group showed significant negative linear relationships for both tests (4028; p's = .10, .03, range = .59 to 1.86 minutes). This variable probably reflects the teacher's management ability, because the time in this component was spent organizing materials and getting students' attention. A well-organized teacher would have the materials ready, and could get student attention quickly.

These three components were also examined as proportions of the total transition time. There were no significant relationships for any of them for the Total Reading score, but the proportion of transition time that was time-to-group for the teacher showed a significant interaction in the Word Analysis score data (4060; $\underline{p} = .03$, range = .04 to .33). Relationships were positive for high and negative for low classes. The negative slope for lower ability classes makes intuitive sense, in that teachers who were with their students sooner were decreasing "dead" time in the group when the students were left without direction. However, the positive findings for higher ability classes are puzzling. In the absence of significant results for the Total Reading data, this finding will not be interpreted at this point.

One of the measures of the effectiveness of attention getters during transitions was a rating on a 5-point scale of the percent of children paying attention to signals when they were given. This average rating showed significant positive linear relationships with both tests (5303; $\underline{p}$'s = .03, .03, range = 3.05 to 3.67 on a 5-point scale). Most teachers had 60 to 80 percent of their students attending to the transition signals, but those who got higher attention were producing higher achievement. Again, this result is probably a reflection of better management by the more effective teachers. However, there were no significant relationships for the average rating of the percent of children attending to signals given at the beginning of the lesson (5304). Prelesson signals were not used very often, and when they were and could be rated, the range was fairly restricted. Therefore, the lack of significance for this variable probably reflects its poor distribution and low frequency of occurrence.

Another measure taken to indicate effectiveness of signals was the number of times teachers repeated signals to start the transition or to get attention in the group (4041, 4043). Both variables had restricted ranges and there were no significant findings.

Another indication of effectiveness of signals was the average number of individual corrections delivered by the teacher during the transition or at the beginning of the lesson. There were no significant results for the average number of corrections during transitions (4042), but there was a significant negative linear relationship for the average number of individual corrections given before starting the lesson (4044; $\underline{p}$'s = .05, .01, range = .17 to .75). The more times the teacher corrected individuals before beginning the lesson, the lower the achievement for the class. This variable probably also reflects the teacher's overall management ability.

Seating the group. The teacher's adherence to Principle 2 (arranging the group) was measured by two ratings. The appropriateness of teacher seating was defined as the number of students who could be seen easily by the teacher (a rating of $\underline{5}$ represented 100%). Likewise, appropriate child seating was defined as the number of other students who could be seen easily by the children in the reading group (where $\underline{5}$ represented 0%). There was a significant positive linear relationship for the average rating of teacher seating (5301; $\underline{p}$'s = .01, .03, range = 2.72 to 4.21). Teachers who arranged themselves so as to see most of the out-of-group students were also those who produced greater achievement. This variable probably also reflects the teacher's overall management ability, in that monitoring is an important part of management. There were no significant findings for the average rating for child seating (5302).

Summary of results for Principles 1 and 2. The first two principles in the model emphasized the importance of getting and maintaining the attention of the students. Some specific suggestions were made in the model for achieving this: using clear signals to call for transitions and to get attention to begin a lesson, and arranging the group so as to make monitoring of the entire classroom easier for the teacher and distractions less likely for the students in the group. Several variables were created to measure both the implementation of these specific suggestions and their immediate effects.

The results suggested that the more effective teachers (in terms of achievement gains) had better control at the beginning of the lesson, as reflected in several short-term outcomes such as time measures and ratings of attention. However, the only specific behavior included in the instructional model that was related to achievement and to these management skills was the teacher seating herself to monitor the out-of-group students. Use of signals (as suggested by the model) did not relate to gain, but there was restricted variance here. (Most teachers in both treatment and control groups used signals for transitions, and very few teachers used them to start lessons.)

However, measures indicating the overall effectiveness of transitions and beginning lessons showed significant positive relationships with gain (e.g., time spent in transitions, rating of student attention, less need for corrections of students once they were in the group). A measure of time spent between the arrival of all persons in the group and the beginning of the lesson showed negative relationships. This could reflect either a lack of teacher preparation and/or more time spent getting students' attention.

In general, then, these findings indicate that the more effective teachers had better control during transitions and when beginning lessons. This is probably due to overall management skills.

Introducing the Lesson and Material to the Students (Principles 3,4,5, & 6)

Using an overview. For each observation, the observer noted whether the teacher gave no overview, whether she gave an overview containing only mechanical content (i.e., the pages to be covered), or whether she gave an overview containing specific instructional content. Proportion variables were created to reflect the number of observations in which the teacher did each of these things. The proportion of time in which no overview of instructional content was given showed a significant negative linear relationship with Total Reading scores (5330; p = .03, range = .38 to .78). However, this variable was not significant for the Word Analysis score.

Several variables described the content of overviews, but none was related to achievement. There were no significant relationships for the variables describing mechanical overviews vs. specific instructional overviews (5331 and 5332). There were also no significant findings for any of the variables describing motivating statements in overviews (5334-5336). The rated enthusiasm of the teacher's voice and of the students also showed no relationship with achievement (5305, 5306). Most overviews were given and received in a neutral manner.

Therefore, the results suggest that teachers who failed to give overviews most of the time had lower achievement, but that the actual content of overviews did not predict achievement.

Presentation of new words. There were no significant relationships for the average number of new words presented in a lesson (5358). There also were no significant results for the proportion of group lessons in which any new words were given or the average number of new words that were given at any one time (5368, 5369).

There also were no significant relationships for the proportion of new words given at the beginning, rather than during the lesson (5359), or for the proportion of new words given by the teacher rather than asked of a child (5360).

However, there were significant findings for the use of clues when new words were presented. There was a positive linear relationship for both tests for the proportion of new words given with phonetic clues (5361; p's = .02, .05, range = .17 to .71). There were no significant findings for the proportion of new words given with context clues or with both phonetic and context clues (5362, 5363). However, there were weak negative findings for the proportion of new words presented with no clues at all (5364; p's = .06, .09, range = .13 to .66). This suggests that teachers who frequently presented new words without any clues at all had lower achievement.

Repetition of new words. When new words were presented, the observers noted whether or not the teachers had the students repeat them and how this was done when it occurred. There were no significant findings for the proportion of new words which were repeated by students (5365). There were no clear-cut findings for the proportion of new words repeated chorally instead of individually, but a weak negative relationship was suggested (5366).

There were no significant findings for the proportion of new words which were repeated by all of the children rather than by some (5367).

In summary of the data describing presentation of new words, no variables related to the treatment demonstrated relationships with achievement. However, using phonetic clues when presenting new words was associated with gain.

Demonstrations and explanations. There were no significant findings for the proportion of activities introduced by a teacher demonstration (5348), or for the proportion not requiring a demonstration (5350). (Most activities were introduced by a teacher demonstration.)

There also were no significant findings for the proportion of demonstrations that had to be repeated because the students did not understand (5351). Other variables measuring effectiveness were 5-point rating scales of the "sufficiency" of the demonstration and the percent of students in the group who apparently comprehended it (as judged by their performance afterwords). There were no significant findings for either of these variables (5310 and 5311). The ranges were restricted, however, which suggests that the ratings may not have been sensitive to variations in quality of demonstration.

The observers noted for each demonstration if and how the teacher checked the students' comprehension. There were five options for this. The proportion of demonstrations in which *student comprehension was checked by having students repeat the instructions* showed no relationships with achievement (5354). However, the proportion of demonstrations in which *student comprehension was checked by having the students actually demon-*

strate the procedures before using them showed significant curvilinear relationships (5355; p's < .01, < .01, range = .06 to .29). For each test score, these curves showed inverted U-shaped relationships for high classes, but shallow positive slopes for low classes. Perhaps this technique is usually appropriate for lower level students who can show their understanding more easily than they can explain it, but it may be inefficient for higher level students if used too much of the time.

The proportion of demonstrations in which student comprehension was checked by starting the lesson in the group so that the teacher could observe the students while they worked showed no significant relationships (5356). Also, the proportion of times that students were sent to their seats after a demonstration without any checks for comprehension showed no relationships (5357). The range of this variable was .19 to .47, indicating that it occurred frequently in some classes but rarely in others.

Summary of results for variables for Principles 3,4,5, & 6. This group of principles made suggestions about presenting new material to students, and was based on the rationale that it is easier for students to receive and process information when it is broken down for them into small chunks. The specific suggestions made were: use overviews at the beginning of the lesson, present new words before they are encountered in reading, have students practice new words when presented, and provide carefully sequenced demonstrations of activities which include checking for comprehension.

Very few of the suggested techniques showed significant relationships with achievement, although some variables which were measured incidentally did show relationships. For Total Reading scores, there was a significant negative relationship for the absence of any overview, but this was not found for the Word Analysis score, and other variables describing the quality or content of the overview showed no relationships. These data suggest that the use of some kind of introductory statement about the lesson may be beneficial, but no conclusions can be inferred about what that overview should include.

There were no significant relationships found for the provision of new words at the beginning of lessons or for the repetition of new words. However, two variables measuring the ways in which these principles were implemented did yield significant results. There were positive relationships for the use of phonetic clues with presentation of new words, and there were weak negative relationships for the absence of any clues at all. These data suggest that presentation of new words should include information about the phonics rules involved.

There were no significant relationships for the use of demonstrations or their rated effectiveness. There were no findings for any of the variables measuring the ways teachers checked for comprehension except for the proportion of times the teacher asked the students to demonstrate the procedures back to her. This variable showed an inverted U-shaped relationship for higher-ability students, but there was more of a straight positive slope for lower ability classes. This suggests that students, especially those of lower ability, are best checked by having them actually show that they know how to do something, rather than by depending on them to ask questions or assuming that they always understand.

There were no relationships for the proportion of time the teacher failed to check students' comprehension of a demonstration before releasing them to do some activity.

Calling on Individual Students in the Group (Principles 7,8,9,10,11, & 12)

Working with one child at a time and giving feedback. The rate of total response opportunities (R.O.'s) given per minute showed significant positive relationships with achievement (601; p's = .08, .05, range = 1.61 to 2.45). The rate of nonturn response opportunities per minute showed strong positive relationships (602; p's = .01, .01, range = 1.00 to 1.84). There were no significant relationships for the number of reading turns per minute or for the number of reading turn response opportunities per minute (603, 604). These results suggest that the more teacher interactions with individual students, the higher the achievement. This was especially true of interactions that were single questions and answers (i.e., nonturn), rather than those occurring within reading turns.

There were no significant relationships for the absolute rate per minute of failures to give feedback following a student's answer (24). Likewise, there were no significant findings for the proportion of total or nonturn response opportunities receiving no feedback (723, 724). The ranges for these two variables were .02 to .15, and .00 to .13, respectively, indicating that omission of feedback did not occur often.

However, when examined only for correct answers, there were weak significant positive linear relationships for the proportion of total correct answers with no feedback (778; p's = .09, .07, range = .00 to .15), and for the proportion of nonturn correct answers receiving no feedback (770; p's = .09, .07, range = .00 to .16).

The proportion of total incorrect answers which received no feedback was not significant (793), but for nonturn R.O.'s with incorrect answers, this variable did show a significant negative linear relationship within a restricted range (794; $p < .01$, $p = .02$, range = .00 to .02).

These results suggest that there is a weak positive relationship for omission of feedback following correct answers, although this did not occur often (less than 15% of the time). However, there were negative relationships for no feedback following incorrect answers, although this was rare. These results are reasonable, even though they are in partial conflict with the treatment principles dealing with giving feedback to student responses. It is obviously inappropriate to leave an incorrect answer without feedback, but a correct answer, especially one that is obviously correct, may not require acknowledgement. Within limits, omitting feedback to correct answers apparently was an efficient strategy.

The frequency of group responses was part of the evaluation of this principle. Two types of group responses were noted. Choral responses occurred when the teacher indicated that she wanted the group to respond in unison. Group call outs were noted when more than one student shouted out in response to a question that was not intended for the group. This distinction was important, because there were different results for the two types of group responses.

There were no significant relationships for the average number of choral responses per observation (4038), the average number of group call outs per observation (4039), or the average number of these combined (4040). However, when choral responses and group call outs were examined as rates of occurrence per minute, several relationships were demonstrated. There were no significant findings for the rate per minute of choral responses and group call outs combined (4053). However, the number of choral responses per minute showed significant negative linear relationships (4051; p's = .05,

.04, range = .04 to .24). The number of *group call outs* per minute demonstrated significant interactions (4052; p's = .05, .04, range = .06 to .34). These showed negative slopes for high level classes and positive slopes for low level classes. These results are similar to those found by Brophy and Evertson (Note 1) with high and low SES classes, in which positive relationships with call outs for low SES classes were interpreted as indications of student enthusiasm and motivation, but negative relationships within higher SES classes were interpreted as indicating that call outs represented control problems in that setting.

These results suggest that it is probably better for teachers to minimize calling for choral responses, but that within some classes, allowing group call outs may be desirable some of the time. However, they should not be so prevalent as to take away time from individual response opportunities, which showed much higher positive relationships, even for lower ability classes.

Methods of selecting students for response opportunities. Five types of initial selection were noted with the coding system. *Initial selection* referred to the way in which the teacher chose a student to answer a question or read aloud at the beginning of what could be a series of interactions. The types of initial selection were:

1. *Ordered*. The teacher selected the students on the basis of their seating position, by moving around the group in a consistent pattern and choosing students in turn.

2. *Preselection*. The teacher called a student's name, and then asked the question. The selection was not based on the order of seating.

3. *Nonvolunteer*. The teacher asked a question, and then called on

a student who had not volunteered to answer (but not on the basis of seating position).

4. Volunteer. These were selections in which the teacher asked a question, and then called on somebody whose hand was up (but not on the basis of seating position).

5. Call out. The teacher asked a question, and before she could call on someone to answer it, another student called out the answer without permission. In order for call out to be coded, the teacher had to respond to the content of the answer by acknowledgement or feedback. If the teacher ignored the student and called on another child, or corrected the student for calling out but did not respond to the content of the answer, the interaction was not coded as a call out selection (i.e., it was not a response opportunity).

There were no significant findings for the rate data describing types of selection (1-7), but several patterns were present in analyses of the proportion data. Proportion variables were created that examined the relative number of interactions chosen by each type of selection. As was done for most proportion variables, these were examined for the separate types of interactions (turn and nonturn) and for total interactions. Response opportunities were also broken down into reading and nonreading questions, reading questions in nonturn interactions, and personal questions. These variables were created to give as precise a picture as possible of the influence of type of selection, and to see if this influence varied according to type of interaction or type of question. Each set of variables is examined below.

The first set are those that are not broken down by type of question, but which are distinguished as to turn and nonturn interactions.

The proportion of all initial selections that were *ordered* was positively linearly related to achievement (607; p's = .04, .05, range = .19 to .77). This positive finding for the use of ordered turns was also demonstrated when interactions were analyzed for reading turns (608; p's = .03, .05, range = .23 to .88), and nonturn interactions (609; p's = .06, .07, range = .17 to .74).

The second category, *preselection*, did not show clear relationships with achievement. For the proportion of total interactions that were chosen by preselections, there was a significant interaction for Total Reading scores, although not for Word Analysis scores (610; p's = .02, .13, range = .00 to .19). The interaction with Total Reading scores showed a negative relationship for high ability classes and a positive slope for low ability classes. The same interactive pattern emerged for both test scores for the proportion of reading turns selected by *preselection* (611; $p < .01$, $p = .03$, range = .00 to .25). Again, the interaction is one in which there is a negative slope for higher ability classes and a positive slope for lower ability classes. When examined for nonturn interactions in which *preselections* were used, there were no significant relationships (612).

The results suggest that preselections are not desirable in higher ability classes but may be useful in low ability classes. However, preselections were not used frequently and so these results must be interpreted in light of the observed ranges: from infrequent to occasional use. Their occasional use in lower ability classes may represent careful matching of question to respondent when this was necessary to catch someone's attention. (That is, the teacher would call the child's name first, and then ask a question.) Greater use might represent teacher sensitivity to student attention and efforts to bring students back into the lesson. This is likely a greater problem with lower ability classes at this grade level, where students may not have developed their attentional skills. On the other hand, the negative results within higher ability classes might represent *too many* problems with inattention, in groups where students should be expected to control their own attention most of the time. (Further analyses at the ability group level within classes may shed more light on the meaning of this interaction.)

The proportion of total interactions answered by *nonvolunteers* showed a significant negative slope (613; p's = .02, .08, range = .08 to .34). Use of *nonvolunteer* selections for reading turns showed a similar pattern (614; $p < .01$, $p < .01$, range = .05 to .39). However, when examined only for selection for nonturn interactions, there were no significant relationships (615). As with other selection variables, stronger relationships with achievement were found for reading turns.

There were no significant relationships for any of the variables describing proportions of interactions selected by volunteers (616, 617, 618). These seldom occurred more than 25% of the time. It had been expected that extensive reliance on volunteering would be negatively related to achievement, due to the more reticent students receiving too little practice. Within the observed ranges, however, this was not supported.

The use of *call outs* as a type of selection was analyzed only for total and nonturn interactions because call outs did not occur very often for reading turns. For each of these, there were significant negative linear relationships: for total interactions (619; p's = .04, .01, range = .03

to .13); and for nonturn interactions (620; p's = .01, < .01, range = .02 to .21). That is, the higher the relative frequency of call outs, the lower the achievement. In the small group setting, call outs may represent a control problem, with a few students attempting to dominate the interactions. If this is so, then the negative findings make sense. Other variables describing call outs suggested that this might be less of a problem in lower ability classes, but this particular variable showed no interactions. It might yield these differences when analyzed at other levels (i.e., ability groups within class) where there would be more variation in ability.

The results described above were not distinguished according to question type. In order to determine if different types of selections might be appropriate for different types of questions (i.e., reading, nonreading, or personal), several variables were created to describe combinations of selection and question type (e.g., the proportion of all nonreading questions selected by ordered turns). It was expected that questions tapping skills should be selected systematically, under careful teacher control, while personal questions might be better selected by volunteering, allowing more student control of who answered.

The same pattern of results was found for these more specific variables as was found for those already discussed. There were no reversals of significant findings. Only two types of selection, ordered and call out, showed significant results for all types of questions examined. All other selection types showed significant relationships only for reading questions.

Teacher responses to call outs. Principle 11 suggested that teachers try to minimize call outs, responding to them with mild reminders that calling out was not acceptable in the classroom. However, it was emphasized

151

that teachers should be careful not to quench the enthusiasm which led to the call out. The relationship of calling out to achievement has been discussed in the previous section and this section examined the ways in which teachers responded to call outs when they occurred.

The proportion of call outs that were selected (that is, were accepted for their content and responded to) and *also* corrected by the teacher was not significantly related to achievement (1032). The distribution for this variable was positively skewed, with a range of .00 to .10. This suggests that most call outs that were accepted were not corrected.

There also were no significant relationships for the proportion of call outs which were corrected but *not* accepted for their content (1001). The distribution for this variable was slightly negatively skewed, with the range being .79 to 1.00. These two variables considered together suggest that teachers who accepted the content of call outs were not likely to correct them, and that if they did correct a call out, they were not likely to accept its content. Therefore, there was no extensive test of the principle of accepting call outs but also gently reminding students that they were not acceptable.

Three types of corrective responses to call outs were examined: *management* (mild corrective statements), *warning* (more severe statements), and *criticism* (most severe, sometimes involving punishment). There were no significant relationships for the proportion of corrected call outs that included *management* (1037) or *criticism* (1039). The ranges for these variables show that management statements were most often used (from 34% to 79% of the time), while criticism was only used rarely (from 0 to 30% of the time that all call outs were corrected).

For the proportion of corrected call outs that included warning statements, there was a significant curvilinear relationship found for Total Reading scores, but not for Word Analysis scores (1038; p for Total Reading = .04, range = .04 to .21). This curve depicts an inverted U-shaped relationship for higher ability classes, and a very shallow relationship for lower ability classes. This suggests that a moderate amount of severity is sometimes the most appropriate correction for students in higher ability classes who call out. However, this was not true for lower ability classes.

No significant relationships were found for similar variables which separately examined corrections to accepted and unaccepted call outs (1020-1024).

Use of comments. Student comments on classmates' responses were not used very often by any teachers, and within the range available for analysis, there were no significant relationships for the proportion of response opportunities receiving comments for Total Reading scores. However, there was a significant interaction for this variable with Word Analysis scores (665; p = .03, range = .000 to .002). This interaction shows a negative slope for higher ability classes and a very shallow positive slope for low ability classes. However, the low frequency of occurrence does not allow meaningful interpretation, and the result probably represents a few teachers using comments very few times.

Use of undesirable types of questions. There were no significant relationships with either criterion for the average number per observation of each type of undesirable question: rhetorical (4045), answering one's own question (4046), asking a series of questions without stopping for answers (4047), miscellaneous undesirable questions (4048), or the total number of undesirable questions (4049). There also were no significant relationships for the average rate per minute of the total number of undesirable questions (4048).

There were significant curvilinear relationships with both achievement scores for the variable measuring the rate per minute that the teacher answered her own questions without waiting for student answers (4055; p's = .06, .04, range = .00 to .02). This range is fairly restricted, and the full distribution is positively skewed. Therefore, this result will not be interpreted, especially in view of the other results for undesirable questions. It must be concluded that the hypotheses about undesirable types of questions were not tested, because so few of them appeared.

Summary of results for Principles 7,8,9,10,11, & 12). These principles dealt with calling on individual students in the group, and made suggestions that individual students be given practice and feedback about important skills; that choral and group responses be minimized; that students should generally be selected in order of seating with minimal volunteering and call outs; that call outs should be corrected, although not harshly; that students occasionally be asked to comment on another student's answer; and that confusing questions should be avoided.

There were strong relationships with achievement for some of the variables measuring these principles, and weak or no relationships for others. The overall rate of interactions with the teacher, especially in nonturn situations, was positively related to achievement. That is, the more questions asked of students, the more they learned.

There were unexpected findings for omission of feedback. There were weak positive relationships with achievement for the absence of feedback following correct answers, although this occurred less than 15% of the time. Only a small proportion of incorrect answers received no feedback, but within this range, there were strong negative relationships with achievement. Therefore, it can be concluded that omission of feedback after correct answers may sometimes be appropriate, but essentially all incorrect answers should receive feedback from the teacher.

The rate of choral responses was negatively related to achievement, although the rate of group call outs showed an interaction (a negative relationship in higher ability classes and a positive relationship in lower ability classes). These results suggest that choral responses should be minimized in all classes, but that occasionally allowing spontaneous group call outs may be a useful technique in some classes, especially in classes of lower than average readiness.

The data on types of initial selection showed very strong findings, in that the use of ordered turns was *positively* related to achievement, while use of nonvolunteers and use of call outs showed *negative* relationships. There were few relationships for use of preselections, although there was an interactive pattern suggesting that high use of preselection was negatively correlated for higher ability classes, but was weakly positively related to achievement in lower ability classes.

Several variables examined the types of corrections delivered to call outs. Very few call outs that were accepted for their academic content were also corrected for the calling out behavior. Data describing corrections showed few significant relationships with achievement, but there was one result suggesting that moderately severe corrections are sometimes appropriate for higher ability classes.

Requesting students to comment on one anothers' answers occurred rarely, and the data are therefore not interpreted.

The use of confusing questions was also rare and there were few significant results.

Responding to Individual Differences within the Group Setting (Principles 13, 14, 15, & 16)

Breaking up the group due to ability differences. There were no relationships with either achievement score for the proportion of times that reading groups were broken up due to ability differences as suggested by Principles 13 and 14 (5338-5340). The ranges indicate that this occurred generally less than 5% of the time. The proportion of time that the observers judged that the group did not need to be broken up ranged from .84 to .98, indicating that the technique recommended by the instructional model would not have been appropriate most of the time. However, even the data from occasions when the observers judged that it was appropriate but was not used yielded no significant relationships. Therefore, this technique was not supported by the data as being useful in this setting.

Using a model within the group. No teachers were observed using this technique. Possible reasons for this are discussed in the preceding chapter.

Arranging for tutorial help. No direct observational data were collected to measure implementation of this principle. It will be discussed in future reports dealing with teacher interview data and individual student data.

Summary of results for Principles related to responding to individual differences within the group setting. No relationships with achievement were found for any of the variables measuring these principles. There was very poor implementation of this part of the model, yielding little or no variance on most of the measures.

Responding to Incorrect Answers (Principles 17,18, & 19)

Use of terminal and sustaining feedback. Whenever a student did not answer a question correctly, the teacher had the option of providing terminal feedback or sustaining feedback. Therefore, a variable was computed expressing the ratio of terminal feedback to the total of terminal plus sustaining feedback. The higher this score, the more the teacher chose terminal feedback as a response to incorrect answers. There were significant negative linear relationships for total response opportunities (757; p's = .01, .02, range = .40 to .66) and for nonturn response opportunities (759; p's = .06, .07, range = .32 to .56). However it was not significant for turn response opportunities (758; range = .43 to .79).

These ranges indicated that all teachers used terminal feedback some of the time, but the results suggest that too much use of it was related to lower achievement. Indeed, teachers with higher achievement scores used more sustaining than terminal feedback.

It is not surprising that there were different results for interactions occurring within reading turns and those outside of turns, because of the pacing requirements of these different situations. Within reading turns, it is important to maintain a steady pace of word-calling in order to understand the meaning of the text. Too frequent use of sustaining feedback in response to errors in turns would probably destroy the pace that is necessary to understand the material. However, all teachers did use sustaining feedback some of the time during turns, and the lack of a significant

157

relationship with achievement for this variable suggests that this was not detrimental all the time. Probably, sustaining and terminal feedback are each appropriate some of the time in response to errors in reading turns, depending on the content being read and the type of error. Unfortunately, context was not described at that detailed a level.

On the other hand, the use of sustaining feedback would not seriously interfere with the pace in most nonturn interactions, with the exception of fast-paced drills, which were not observed very often. In most nonturn interactions, the teacher asked a single question for the purpose of teaching a skill and allowing practice with it. When errors were made, the use of sustaining feedback, especially in the form of clues, may have provided further information about the skill being taught, and it was given at the time that such information was needed (i.e., after an error). The pace of such lessons was not as important as that of reading turns, where stopping to work through some decoding skill would probably mean loss of practice of another important skill; comprehending the passage.

Types of terminal feedback. In order to examine the effects of different types of feedback more closely, the separate categories were analyzed in a variety of ways. First of all, each type of terminal feedback was expressed as the proportion of response opportunities which included it.

For total interactions, the proportion that included *give answer* feedback was related to achievement, although this was curvilinear for Total Reading scores and linear for Word Analysis scores. For the Total Reading scores, the curve was an inverted U-shaped for both high and low ability classes. For the lower level classes, however, the relationship was more clearly negative for scores at the upper end of the range (735; $p = .05$, range = .04 to .19). For Word Analysis scores, there was a significant negative linear relationship for the proportion of total response opportunities that included *give answer* feedback (735; $p = .01$, range = .04 to .19).

Therefore, when expressed as a proportion of all interaction, more extensive use of *give answer* feedback was related to lower achievement. However, remember that this variable included all response opportunities, both those which were correct and required no feedback, and those for which sustaining feedback was more appropriate. The range indicates that teachers at the high end on this variable were giving the answer in almost one out of five interactions. This might reflect too high an error rate, and it might also reflect an unwillingness to use sustaining feedback to errors. That is, giving the answer is not necessarily an undesirable technique (the more effective teachers used it some of the time), but this result suggests that it can be used inappropriately.

There were no significant relationships with achievement for the proportion of turn (736) and nonturn (737) response opportunities that contained *give answer* feedback.

The proportion of response opportunities containing *ask other* feedback was not significantly related to gain for either total interactions (738) or nonturn interactions (739). The range of scores for this variable was limited (from .01 to .09). It had been expected that this would be negatively related to achievement, especially in lower ability classes.

The occurrence of *call out* feedback was examined in the same way. The proportion of total response opportunities that contained *call out* feedback showed different results for the two test scores. With Total Reading scores, there were no significant results. For Word Analysis scores, there was a significant curvilinear U-shaped relationship that was generally negative in slope (740; p = .01, range = .00 to .03). The restricted range demonstrates that calling out as feedback was infrequent when compared to all interactions. There were no significant relationships with either achievement score for the proportion of *turn* interactions which involved *call out* feedback (741). However, when *nonturn* interactions were examined, significant linear interactions were found for both test scores (742; p = .03, .01, range = .00 to .02). The interactions depicted a negative linear relationship for higher ability classes, and a very shallow positive one for lower ability classes. This same interactive pattern has been demonstrated elsewhere in this study and in others (Brophy and Evertson, Note 1). Such interactions are interpreted to mean that calling out behavior (within limits) *may* indicate enthusiastic participation in lower ability classes, but control problems in higher ability classes where motivation is not as much of a problem.

To summarize the results for the types of terminal feedback (expressed as proportions of all interactions), negative relationships were found for the use of *giving the answer* to students too frequently, and also for higher levels of *call outs* from students in higher ability classes. There was a positive relationship (within a very small range) for *call outs* as feedback within lower ability classes, at least for nonturn interactions. There were no significant relationships for the use of *asking another student* for the answer.

In order to examine the single types of terminal feedback in more detail, variables were created that expressed each specific type as the proportion of all terminal feedback that was used. These variables tell us the relative importance of the three specific techniques when terminal feedback is selected as a general strategy. That is, if the teacher is not going to give sustaining feedback, is one type of terminal feedback more closely related to achievement than another? Is the appropriate use of terminal feedback dependent on the specific kind of feedback used?

In the instructional model, *giving the answer* to students was considered better than *asking another student* for the answer or having students *call out* feedback without permission. Therefore, positive relationships were expected for the relative use of *giving the answer*, and negative results for the other two techniques. However, the proportion of terminal feedback that was *give answer* feedback showed no significant relationships with either achievement test (760, 761, and 762). The ranges for this measure indicated that *give answer* feedback was used frequently when terminal feedback was selected as a general strategy. For total response opportunities, the range was .49 to .85; for turn interactions, it was .74 to 1.00, and for nonturn interactions it was .23 to .56.

This technique was used more frequently than other types of terminal feedback during reading turn interactions, and up to half the time that terminal feedback was used in nonturn interactions. Again, the more frequent relative use of *giving the answer* to errors in reading turns probably reflects the pacing requirements of that setting. *Giving the answer* takes less time than the other types of terminal feedback, and this is an important consideration in reading turns.

The proportion of terminal feedback that was *ask other* showed significant curvilinear relationships for total response opportunities (763), but was not significant for nonturn interactions (764). This variable was not examined separately for turn interactions, because of low frequency. The curves for total interactions were inverted U-shaped, although much more shallow for Word Analysis scores than for Total Reading scores (p's = .01, .04, range = .08 to .39). This suggests that moderate use of this type of feedback is appropriate when terminal feedback is being given, at least in nonturn interactions.

Earlier research which examined *ask other* feedback demonstrated different relationships for SES groups: it was positively related to achievement for high SES classes but negatively related for low SES classes (Brophy and Evertson, 1976; Note 1). The curvilinear result does suggest that too much use of this technique would not be advisable, but that it is not totally undesirable, at least in nonturn interactions. The low rate of occurrence in reading turns suggests that most teachers find it inappropriate then, probably because of the effect on pacing.

The proportion of terminal feedback that was a *call out* by another child showed no significant relationship for either test score (765, 766, 767).

These results do not yield any clear conclusions about the most appropriate type of terminal feedback. The appropriate use of any type of feedback is probably dependent on several factors, such as pacing requirements, student characteristics, and the extent of use.

Types of sustaining feedback. The three separate categories of sustaining feedback also were examined in the same ways to determine more about their effect on achievement.

There were no significant relationships for the proportion of total response opportunities that included repeating the question feedback (746). For turn response opportunities, there was no relationship with Total Reading scores, but there was a significant linear interaction for Word Analysis scores (747; $p = .04$, range = .04 to .19). In this interaction, there was a positive slope for higher ability classes, and a negative slope for lower ability classes. There were no significant relationships with either test for the proportion of nonturn response opportunities which included repeat question feedback (748; range = .03 to .06). The ranges indicate that repeating the question as feedback was used most often in reading turns and seldom in nonturn interactions. It may be that it is more effective with higher level students because they are more capable of figuring out for themselves what should be done to correct a misread word, and this is an efficient use of time within reading turns. Perhaps lower ability students need more information than is provided by simply repeating a question after a mistake. This technique might be effective when an error was caused by impulsive guessing or reading too quickly, but it would be less effective if errors were due to the lack of skills required to decode a word. If the latter type of error were more common in lower ability classes, then it makes sense that too frequent use of repeating the question would not be useful, because it would not provide the information which the child needed to correctly read the word.

The proportion of response opportunities which contained clue feedback showed no significant relationships for either total, turn, or nonturn interactions (749, 750, 751). The range here was small, from .06 to .10, indicating that few interactions included such feedback.

There also were no significant findings for the proportion of any type of interaction that was *give by clue* feedback (752, 753, 754). The range was restricted on this measure, from .00 to .01, reflecting the low level of use of this technique.

The next set of variables expressed the proportionate use of each of these three techniques to one another.

The proportion of all sustaining feedback that was *repeating the question* yielded no significant relationships for total R.O.'s (768). However, for turn response opportunities, there was a linear interaction with a positive slope for higher ability classes and a negative slope for lower ability classes (769; p = .09, .01, range = .24 to .53). This is the same pattern which was demonstrated when *repeating the question* was expressed as a proportion of all response opportunities, and the interpretation is the same (i.e., lower ability students may need more information, whereas higher ability students can more often benefit from having the question repeated). There were no significant findings for *repeating the question* in nonturn interactions (770).

Although no significant findings resulted when *clue* feedback was expressed as the proportion of all response opportunities, there were significant interactive results when it was examined as the proportion of all sustaining feedback in total and in turn response opportunities.

For total response opportunities, the interaction was a negative relationship for higher ability students and a positive relationship for low ability students (771; p = .10, .04, range = .47 to .70). For turn interactions, this same pattern was present and the results were more highly significant (772; p = .01, $p < .01$, range = .43 to .70). For nonturn interactions with sustaining feedback, there was no significant relationship (773).

These results suggest that when a teacher decides to use sustaining feedback within reading turn interactions, it is probably better not to use clue feedback too much of the time with higher level students. This result complements the ones discussed earlier for repeating the question. Higher level students probably do not need as much information to correct their own answer, at least within reading turns, and providing too much of it probably breaks the pace. The ranges for these scores indicate that all teachers gave clue feedback some of the time when they used sustaining feedback in turns, so this result should not be interpreted to mean that clue feedback should never be used with higher level students. It simply shows that too much of it yielded negative relationships. This was not true with lower ability students, who apparently benefited from more clue feedback in reading turns compared to other types of sustaining feedback. The implication is that when teachers want a student of lower ability to try to correct an error while reading (that is, they have decided on sustaining feedback rather than terminal), it will probably be better to offer clues instead of simply urging the child to try again.

The proportion of total sustaining feedback that was given by clue feedback showed no significant relationships with either test (774, 775, 776). Give by clue feedback was used from 2% to 9% of the time that sustaining feedback was used.

In summary of the data describing the specific types of sustaining feedback, no significant relationships were found for the use of give by clue feedback when examined in these ways. Repeat question feedback and clue feedback yielded significant interactions with ability when examined for reading turns. Higher ability students making errors in turns who received more repeat question and less clue feedback had higher achievement, while these relationships were the opposite for lower ability classes.

Feedback given when students fail to respond. The next three sections will look at feedback categories separately for three types of answers: failure to respond, incorrect answers, and statements of "I don't know".

This first section examined what kind of feedback is given when students fail to respond. Therefore, the denominator for these proportions was "the sum of all no response answers" instead of "all interactions".

The proportion of all no response answers which included give answer feedback demonstrated significant negative linear relationships with both test scores (833; p's = .01, .02, range = .19 to .55). The more often the teacher gave the answer to a child who had failed to respond initially to a question or word, the worse was the overall achievement. This does not mean that giving the answer is never an appropriate thing to do, however, because the bottom of the range represented teachers who did so almost 20% of the time that students did not respond. When this variable was examined separately for turn and nonturn interactions, no significant relationships were found (834, 835).

The proportion of no response answers followed by ask other feedback showed no significant relationships for either total or nonturn response opportunities (836, 837).

The proportion of no response answers which included a call out from another child as feedback showed no significant relationships for total interactions or turn interactions (838, 839). However, there were significant interactions for the proportion of no response answers in nonturn interactions which included a child *calling out* the answer (840; *p* < .01, < .01, range = .01 to .16). This interaction includes a negative slope for higher ability classes and a positive slope for lower ability classes. This suggests again that call outs are not desirable in higher ability classes (possibly because they represent management problems there), but they may represent something very different in lower ability classes. However, this result must be considered in light of the observed range, since call outs as feedback did not occur with great frequency. Indeed, other results from the study suggest that too many call outs are not appropriate. These interactions with ability suggest that a slightly higher level of call outs should be tolerated in lower ability classes because they may represent something desirable (motivation, enthusiasm, attention to the lesson). This does not mean that teachers of lower ability classes should encourage *high* levels of calling out among their students.

The proportion of no response answers which included *repeat question* feedback showed no significant findings for total response opportunities (844), but did yield a significant interaction for turn response opportunities (845; *p*'s = .08, .02, range = .00 to .14). This interaction showed a positive slope for higher ability classes. This further substantiates the pattern found when *repeating the question* was examined with respect to all answers and all sustaining feedback. There were no significant findings for nonturn interactions including no response answers followed by *repeating the question* (846).

No response answers followed by *clue* feedback also yielded significant interactions, as well as overall linear relationships with achievement.

For total response opportunities, there was a significant interaction with a steep positive slope for higher ability classes and less steep positive slope for lower ability classes, indicating that this technique was more closely related to achievement in higher ability classes (847; p's for interaction effect = .03, .04, range = .17 to .41). When the noninteractive model was considered, there was still a significant positive relationship (p's = .01, .01). This variable was not significant in separate examinations of turn and nonturn interactions, although the results were near-significant and the slopes were positive.

The proportion of no response answers (in total interactions) that were followed by give by clue feedback showed a curvilinear relationship with the Total Reading test, but no significant relationship with the Word Analysis test. The curve essentially described a positive relationship (850; p = .01, range = .00 to .06). Obviously, this technique was not used very often following failures to respond. When this variable was examined for turn interactions, there were significant linear interactions for both tests (851; p's = .04, .04, range = .00 to .04), with positive relationships for both high and low classes, but much steeper slopes for lower ability classes. This suggests that, within reading turns, using a very easy clue question after a failure to respond is occasionally advantageous, especially in lower ability classes. There were no significant findings for this variable for nonturn interactions (852).

In summary, the following conclusions may be drawn from the data describing feedback to failures to respond:

1. Giving the answer to students without requiring another response has a negative relationship with achievement. It may be that doing this

too frequently reinforces the failure to respond, and this may cause students to give up too easily, therefore getting too little practice in figuring out how to read. That is, if students know they can get the answer from the teacher without trying, they may be more likely to stop prematurely and ask for help.

2. The relationship with achievement for other feedback techniques depended on the ability level of the class. The presence of some calling out by other students within a limited range was positively related to achievement for lower ability classes, but this relationship was negative for higher ability classes. This pattern was demonstrated with some other variables in the study describing calling out and in previous research. Clueing feedback (i.e., breaking down the initial question into a series of simpler questions, and requiring another response) was positively related to achievement for both high and low ability classes, but the strength of the relationships within each ability group depended on the type of clue given. For regular clues, when the answer to the clue question was not an obvious "giveaway", higher level classes showed steeper positive slopes than did lower level classes. However, when much simpler clues were examined, there were steeper relationships for lower ability classes, at least within reading turn interactions.

Another interaction with ability level was present for repetition of the question after a failure to respond. Within reading turns, a failure to respond meant that the student stopped while reading the text and repeating the question meant that the teacher simply prompted the student and encouraged an attempt. There were positive relationships with achievement in higher ability classes, but negative relationships in lower ability classes.

These results suggest that optimal response by the teacher to a student failure to answer a question will vary depending on the ability level of the student. For higher ability students, the optimal response seems to be one that is somewhat challenging: pushing for an attempt to read the word within a turn, offering clues that are not too easy, discouraging call outs from other students who would give away the answer, and minimal use of giving the answer to the student by the teacher.

However, for students in lower ability classes, the results suggest that feedback to failures to respond should not be overly challenging, but should not reinforce the nonresponding behavior. Giving the answer to the student by the teacher should therefore be minimized, in order to reinforce the expectation that students should respond. Allowing occasional call outs from other students who provide the answer is not inappropriate. Giving clues is often appropriate, but at least some of the time, those clues should be very simple, allowing the student a guaranteed correct answer. This appears to be especially important in reading turns, where pacing must be considered.

Such interpretations are in keeping with other research that has indicated that higher ability students in the early grades benefit from relatively more challenge, while lower ability students benefit from a relatively higher success rate and more encouragement (Brophy and Evertson, 1976; Note 1).

Feedback given to incorrect answers. In addition to examining these six types of feedback, the variables describing response to incorrect answers also included failure to provide feedback (i.e., there was not even acknowledgement of incorrectness).

The proportion of incorrect answers that were not followed by any feedback showed no significant relationships for total response opportunities (793), probably due to the very restricted range of scores (from .00 to .01). However, when examined for nonturn interactions, there was a significant negative linear relationship (794; $p < .01$, $p = .02$, range = .00 to .02). Obviously, this did not happen very much of the time, but those teachers who occasionally did fail to give feedback to incorrect answers had lower achievement scores. This seems obvious, since students in this situation were essentially receiving inaccurate information about reading, and/or the message that their performance was not important.

The proportion of incorrect answers (total interactions) that received *give answer* feedback showed a significant relationship for Total Reading scores but not for Word Analysis scores. With Total Reading there was a significant curvilinear relationship, with both low and high ability groups showing inverted U-shaped curves. However, for the low ability group, the curve dropped off much more steeply after peaking. For the high group, it leveled off more gradually (803; $p = .02$, range = .21 to .49). This suggests that high levels of *giving the answer* to incorrect responses is especially undesirable for lower level students. However, no significant relationship was found for the Word Analysis test for total response opportunities, nor were there significant results for turn and nonturn interactions for either test score (804, 805). Therefore, this interpretation remains tentative.

The proportion of incorrect answers that received *ask other* feedback showed no significant relationships for either test for either total or nonturn interactions (806, 807).

The proportion of incorrect answers that received call out feedback for total interactions showed significant curvilinear relationships (808; p's = .10, .01, range = .00 to .06). The curve for lower ability classes suggested a weak positive relationship. For the high classes, the curve was primarily negative in slope with a short upward curve at the end. Therefore, even though curvilinear, the pattern is essentially the interaction detected earlier for call outs: negative relationships in higher ability classes and positive relationships in lower ability classes. This relationship was not found when separate analyses were done for turn and nonturn interactions (809, 810).

The proportion of incorrect answers that received repeat question feedback showed a significant curvilinear relationship (inverted U-shaped) for total response opportunities for the Total Reading test (814; p = .06, range = .12 to .30). For the Word Analysis test this variable showed a significant linear interaction (814; p = .04, range = .12 to .30) with a positive slope for high classes and a negative slope for low classes. Separate analysis of this variable for turn and nonturn response opportunities yielded no significant results (815, 816).

For total response opportunities, the proportion of incorrect answers that received clue feedback showed no significant findings, although there were near-significant results indicating a positive linear slope (817; range = .15 to .33). When turn and nonturn interactions were examined separately, no significant relationships were found (818, 819).

The proportion of incorrect answers which received give by clue feedback showed no significant relationships with gain for total interactions, turn, or nonturn interactions (820, 821, 822).

Incorrect answers were divided into reading questions and nonreading questions to determine if this finer analysis would reveal a different pattern of findings. That is, would the appropriateness of feedback vary according to these types of questions? The following patterns were evident:

1. There were no significant relationships with achievement for any of the variables including nonreading questions. These were examined for nonturn interactions only, since so few nonreading questions occurred in turns.

2. When reading questions were examined separately, a similar pattern of results was found to that already discussed for all questions that yielded incorrect answers. Only the results that were different are discussed below.

There were no significant relationships for giving the answer to incorrect reading questions (888, 890, 891), whereas analyses involving all questions, yielded a curvilinear (inverted U-shaped) relationship for total R.O.'s with one test.

A significant negative linear relationship was found for the use of asking other students for the answer in nonturn reading questions (892; p = .02, .03, range = .03 to .31). In contrast, there were no significant findings for asking others after incorrect answers for all questions combined.

More significant findings were evident for variables describing repeat question feedback. When all questions were combined, there were significant relationships only for total response opportunities, but for reading questions, there were significant results for all three types of interactions. For total and turn response opportunities, relationships were curvilinear, inverted U-shaped, and similar for both ability levels. For nonturn R.O.'s, the curves suggested an essentially positive relationship for higher ability classes, with no clear relationship for lower ability classes.

These results suggest that moderate amounts of *repeating question* feedback are appropriate, but that the optimal level is dependent on the ability level of the students. This same pattern has been discussed elsewhere for other variables describing *repeating the question*. The interactions with ability are probably due to differing needs for information in order to correct an answer.

Weak positive linear relationships were found for use of *clue* feedback following incorrect answers to reading questions (total response opportunities) (902; p's = .10, .07; .15 to .36). This variable was not significant for all questions combined.

In summary, the following conclusions may be drawn from the data describing feedback to *incorrect* answers:

1. Many of the same patterns found for feedback to *no response* answers were evident, but they were not as consistent or as strong. This suggests that, as expected, a failure to respond and a response that is incorrect require different things of the teacher. In the first case, the task is relatively clear cut; to encourage the student to say something. (This point was strongly emphasized in the instructional model.) However, when the student has answered incorrectly, the teacher's options are more likely to be defined in terms of the error itself and its importance in the lesson at that point. Therefore, the object becomes providing the correct information in the best way.

2. Giving the answer to the student, which showed strong negative relationships with achievement for *no response* answers, was only weakly

related to achievement following incorrect answers. This seems reasonable, in that too frequently giving the answer after a failure to respond may indeed reinforce that behavior, but giving the answer after an incorrect answer may be the most efficient way of providing the correct information. However, the data do suggest that too much use of the technique following incorrect answers is not appropriate.

3. There were interactions with ability for the use of repeat question and call out feedback similar to those already described (higher ability classes had higher achievement with fewer call outs and more repeating the question, and the reverse was true for lower ability classes). However, again, these were not as strong as they were when examining failures to respond.

4. The strength of the relationship with achievement depended on the type of question. There were no significant findings when nonreading questions were examined separately. However, when reading questions were examined, a similar pattern of results emerged to that already described, and for some variables the results were stronger. Of particular importance here was a negative relationship with achievement for asking another student following an incorrect answer. This is the only instance of a negative finding for this technique, although the original hypothesis was that it would be negatively related to achievement.

Feedback given to "I don't know" answers. The same types of feedback variables were created to express teacher responses to answers of "I don't know" or requests for help. (In the model, there was much emphasis placed on eliciting some kind of response to every question, with the understanding that it is often appropriate for the child to say "I don't know", although it is not appropriate to say nothing.)

However, "I don't know" answers did not occur very often, so that many of the feedback categories associated with them were not suitable for analyses due to low frequency.

There were no significant results for use of give answer feedback (823, 824, 825), ask other feedback (826, 827), or repeat question feedback (828, 829). Calling out by other students and giving the answer by a clue were not examined for "don't know" answers due to low frequency.

The only significant finding for the cluster of variables describing feedback to "I don't know" answers was for the use of clue feedback in nonturn response opportunities. Here, the relationships were curvilinear (832; *p*'s = .04, .05, range = .10 to .47). For both ability groups, these curves showed essentially positive relationships with a plateau for high levels of the variable. This suggests that giving a clue when students admit they don't know the answer is an appropriate thing to do, at least within the range of that behavior that was observed.

Results of sustaining feedback. In addition to looking at relationships between achievement and types of feedback in different situations, the effects of sustaining feedback also were examined. Every time that the teacher gave sustaining feedback, the observer noted whether or not it led to an improved response in the next interaction. Improvement was defined as any correct answer or, in the case of an initial failure to respond, as making any response. Proportions were then created to express the number of times that such feedback led to improvement.

The proportion of all interactions involving sustaining feedback that resulted in improved answers yielded no significant relationships for either total, turn, or nonturn interactions (934, 935, 936). It had been expected that teachers who got better results with sustaining feedback would be the more effective teachers, but this was not the case. The ranges on these variables showed that sustaining feedback usually resulted in improvement, although there was still variation among the teachers. For example, for total response opportunities including sustaining feedback, teachers' improvement scores ranged from 63% to 77% of the time. Therefore, they were successful most of the time in eliciting a better answer in the next interaction, although this was not necessarily the final correct answer. Therefore, these measures did not define whether or not the teacher led the student to the final solution. It might be that a measure of ultimate success or failure with a sequence of sustaining feedback would be a better indication of the teacher's ability to use it effectively, and this in turn might be related to achievement. It seems likely that teachers who are able to accurately diagnose the cause of an error and lead the child successfully through the process of answering the question would be more effective teachers overall. However, this hypothesis was not tested by these particular variables.

Each type of answer (incorrect, "I don't know", and failure to respond) was also examined for improvement following it. This was done in two different ways. Scores were created which expressed the improvement ratio for only those interactions which were sustained (i.e., those which could be completely categorized as either improved or not improved). Also created were variables expressing the proportion of improved answers out of all such answers (i.e., those which included sustaining feedback, as well as all the times that the answer was not sustained).

For the first type of variable, no relationships were found for the proportion of *sustained incorrect* or *no response* answers that were improved (928, 942). However, when "I don't know" answers were considered, there was a significant interaction for the proportion of these which were sustained and improved in nonturn response opportunities (940; p = .02, .06, range = .43 to .76). This interaction showed relationships with achievement which were positive for higher ability classes, and negative for lower ability classes. It is not clear why there was a negative finding for low classes, especially in light of other findings that suggested that clues which guaranteed improvement were positively related to achievement in these classes.

When the proportion of each type of answer that was improved was examined, positive relationships with achievement were found for incorrect answers (937; p's = .06, .09, range = .23 to .43). Since these variables were examining the proportion of *effective sustaining* feedback to all other types of feedback, and since most sustaining feedback was effective, these results may actually reflect the general appropriateness of using sustaining feedback much of the time. However, there were no significant findings for this variable for "I don't know" answers (939). Again, results for "I don't know" answers do not fit into a pattern established for other types of answers. This might be due to their low frequency, which could lead to unreliable results, or it might be that they actually represent a qualitatively different situation than incorrect answers and "no response" answers.

Each type of sustaining feedback was examined for the proportion that resulted in improved responses. No significant relationships with achievement were found for any of the three categories (944 - 951). Not surprisingly, give by clue feedback more often resulted in success when it was used. but relative success with this or any other type of feedback did not predict achievement. These support the more general variables already discussed which showed that effective use of sustaining feedback as it was measured did not relate to achievement.

Another set of variables was created to examine the relative length of a series of sustained interactions, in order to determine if teachers who gave up before ultimate success differed from teachers who usually continued sustaining sequences through to the end. The variables expressed the proportion of sustained interactions which did not include the final correct answer that were terminated rather than sustained further. It might be predicted that the relationship of this variable to achievement would be curvilinear, in that giving up and offering terminal feedback too much of the time might communicate negative expectations, while never being willing to stop before success could seriously disrupt the pace of the lesson. However, there were no significant relationships for this variable for total, turn, or nonturn interactions (662, 663, 664). Values ranged from 30% to 50% for sustained interactions receiving terminal rather than further sustaining feedback.

Summary of results for Principles describing feedback to incorrect answers (Principles 17,18, & 19). The principles in the model suggested using sustaining feedback in response to errors when it seemed appropriate. When it was not appropriate (due to pacing or type of question), it was suggested that the teacher give the answer to the student herself rather than ask other students.

In a very general sense, these principles were supported by the data, but the results suggested that many factors must be considered in defining their relationship with achievement. There were several statistical interactions with the initial readiness level of the class, and results were often different when separate analyses were done for response opportunities during reading turns and those outside of reading turns.

Sustaining feedback was positively related to achievement, when examined in comparison to terminal feedback. However, the results indicated that terminal feedback was used by all teachers much of the time. The differences between the more and less effective teachers was in the relative amount, with the more effective teachers using sustaining feedback up to 60% of the time that they had the choice of the two general strategies. In addition, the proportion of incorrect answers and failures to respond that were improved through the use of sustaining feedback were positively related to achievement.

When specific types of feedback were examined, the relationships with achievement were dependent on both student ability level and the type of answer preceding the feedback.

Terminal feedback categories:

Giving the answer showed negative relationships with achievement, and this result was strongest when failures to respond were examined. However, all teachers, even the most effective, used the technique some of the time. The results therefore suggest that it should be used in moderation.

Asking other students for the answer did not occur very often, and there were few relationships with achievement for its use. One significant finding indicated a negative relationship, while another suggested a curvilinear relationship in which moderate use was related to achievement.

This was contrary to expectations based on other research.

Call outs from other students yielded interactions with class mean ability, in that there were positive relationships with achievement within lower ability classes, but negative relationships within higher ability classes. However, there was not much calling out as feedback, so that the positive relationships in low classes do not imply high levels of the behavior.

Sustaining feedback categories:

Repeating the question often showed interactions with ability, in that its use within reading turns was positively associated with achievement for higher ability classes, but there was a negative relationship for lower ability classes. This was interpreted to indicate that lower ability students often need more information in order to improve an answer, but that higher ability students, given some additional time and encouragement, may be able to reason out a word that they originally hesitated on or missed.

Clue feedback following errors or failures to respond showed positive relationships with achievement, although again, this was strongest following failures to respond, and especially in higher ability classes.

Give by clue feedback did not occur often, and yielded few significant relationships. However, its use was positively related to achievement when following failures to respond. This effect was strongest in lower ability classes.

A teacher's rate of success with sustaining feedback (i.e., whether or not it led to an improved answer in the next interaction) did not predict achievement. Most sustaining feedback was successful in this respect.

Feedback to Correct Answers (Principle 20)

There were no significant relationships for the proportion of correct answers followed by *emphasis* feedback for any type of interaction (780, 781, 782, 854, 863, 864, 865, 910, 911). *Emphasis*, as it was measured in this study (repeating the answer or having it repeated), showed no relationships with achievement within the observed ranges. The proportion of correct answers (nonturn interactions) followed by *emphasis* feedback ranged from 26% to 50%, indicating fairly frequent use of this type of feedback. The range was much smaller in reading turn interactions (0 to 14%) probably because the teacher wanted to maintain the pace and repeating words would have been an unnecessary interruption.

It had been expected that omitting feedback after answers would be negatively related to achievement at this level. This was true when incorrect answers were analyzed. However, when the proportion of correct answers that received no feedback was examined, there were weak positive linear relationships for total interactions (778; p's = .09, .07, range = .00 to .15). This variable was not examined in reading turns due to low frequency of occurrence.

When broken down into types of questions, the proportion of correct reading questions that received no feedback showed similar relationships, either weak positive or near-significant trends in that direction (861, 863). However, there were no significant relationships for the proportion of correct nonreading questions that received no feedback (908, 909).

Summary of results for Principle 20 (responding to correct answers). Although it had been expected that *emphasis* feedback (repeating or having a child repeat the answer) would be related to achievement, no relationships were found. Teachers did use this type of feedback fairly often in nonturn interactions (use ranged from one in every four, to one in every two interactions).

It had also been expected that failure to give feedback to correct answers would be negatively related to achievement, since it would represent lack of information given to the student about an answer. However, there were weak positive relationships with achievement found for the proportion of correct answers that did not receive feedback. Omission of feedback to correct answers did not occur often (no more than 15% of correct answers received no feedback). Perhaps omission of feedback is appropriate when it is obvious to the students that an answer is correct, and acknowledgement of this would be unnecessary.

Praise and Criticism (Principles 21 & 22)

Praise. Two types of praise were examined: praise during academic interactions (response opportunities) and praise of behavior. The absolute rate of occurrence of each was not related to achievement (27, 425).

The proportion of all contacts, including response opportunities and behavior contacts, that included praise showed no significant relationships with the Word Analysis test, but significant curvilinear relationships with Total Reading scores (1033; p = .05, range = .04 to .14). The curves for both groups were shallow inverted U-shapes, suggesting an optimal amount of praise in the middle of the range. However, the lack of significance for the Word Analysis test makes this interpretation tentative.

When only academic contacts (response opportunities) were examined, the proportion (total interactions) including praise was significantly negatively related to the Total Reading scores, although not to Word Analysis scores (730; p = .04, range = .04 to .17). There were no significant findings for the proportion of turn response opportunities that included praise, probably due to the very restricted range (731; range = .00 to .03). However, when nonturn interactions were examined, the proportion that received praise showed negative linear relationships with both test scores (732; p = .02, .08, range = .04 to .19).

When only behavioral contacts were examined, the proportion of these with praise showed significant linear interactions with both test scores (1014; p's = .09, .02, range = .00 to .08). These interactions showed a positive slope for higher ability classes and a negative slope for lower ability classes. However, the distribution of this variable was extremely skewed, so that the results may be artifactual. Other research has suggested that lower ability students probably benefit from more praise relative to higher ability students (Brophy and Evertson, 1976; Note 1).

When only correct answers to academic questions were examined, the proportion receiving praise showed no significant relationships to gain for any type of interaction (785, 786, 787, 868, 869, 870).

The use of praise following reading turns that were completely correct showed a negative linear relationship for the Total Reading score (855; p = .05, range = .04 to .30), but this variable was not significantly related to the Word Analysis scores.

The use of praise following incorrect answers showed no significant relationships to achievement for either total or nonturn interactions, probably due to the very restricted range (799, 800; range = .00 to .01).

However, the use of praise following a reading turn that contained some errors showed a significant negative linear relationship with Total Reading scores, although not with Word Analysis scores (859; $\underline{p} = .04$, range = .04 to .33).

Specificity of praise. There were no significant results for the proportion of all praise (academic and behavioral) that was specific (1034). When only academic praise was examined for specificity, there was a positive linear relationship with the Total Reading score, but not with the Word Analysis score (756; $\underline{p} = .04$, range = .01 to .08).

When only behavioral praise was examined, there were no significant relationships to achievement (1029).

Criticism. Criticism was also examined for both academic and behavioral contacts. There were no significant results for the absolute rate of occurrence of either type (28, 429).

The proportion of all contacts (academic and behavioral) that were critical in nature showed significant negative linear relationships with achievement (1035; $\underline{p} < .01$, $\underline{p} < .01$, range = .02 to .10). This variable included academic criticism, behavioral criticism and behavioral warning. It did not include behavior corrections that were mild in tone.

When only academic criticism was considered, there were no relationships with achievement for either total or nonturn interactions (733, 734). Academic criticism was rare, occurring in less than 10% of all response opportunities.

When only incorrect answers receiving criticism were examined, there were also no significant findings (801, 802, 886, 887, 921). Criticism was given to incorrect answers less than 5% of the time.

Specificity of criticism. The proportion of all criticism and corrective contacts that were specific showed positive relationships with achievement (1036; p's = .06, .02, range = .02 to .10). This variable suggests that teachers who were specific about their criticism were getting higher achievement than teachers who were not, at least within the narrow range observed.

However, when only behavioral contacts were examined for use of specificity, there were no significant relationships (1019, 1030). Academic criticism was not examined separately for use of specificity due to the low frequency of occurrence.

Summary of results for Principles 21 & 22 (praise and criticism). The use of academic praise was negatively related to achievement, suggesting that too much praise is inappropriate. These results were strongest for nonturn interactions. For many of the variables describing praise, results were only significant for the Total Reading score, which suggests some caution in interpretation, since most of the other variables were related similarly to both tests.

As expected, use of specific academic praise was related to achievement, but it did not occur often (less than 8% of praise was specific).

There were negative relationships with achievement for the proportion of contacts that were critical in nature. This is a reflection of more critical behavioral contacts, since there were relatively few instances of academic criticism.

When examined separately, there were no significant relationships with achievement for academic criticism.

The proportion of total critical statements that were specific was positively related to achievement.

Time Usage

There were significant relationships with achievement for the average time spent in reading groups. (This included transition time as well as lesson time.) For the Total Reading score, this relationship was positive and linear (4035; $p < .01$, range = 21.24 minutes to 31.84 minutes). For Word Analysis scores, the relationship was curvilinear ($p = .04$). The shape of this curve was similar for both low and high groups, and showed a positive slope up to a plateau.

Similar results were found for the variable "Average response opportunity time available". This variable measured the average time each reading group spent in actual academic lessons, thus eliminating the transition time prior to the lesson. For both test scores, these relationships with achievement were curvilinear (4029; $p = .02, .02$, range = 18.29 to 28.26), and the curves were again positively sloped up to a plateau. For Word Analysis, the curve for the lower ability classes peaked and leveled off much sooner, while the slope for the higher ability classes continued to climb to a higher point before plateauing. These results suggest that the longer students were exposed to instruction in lessons, up to a point, the more they learned. The optimal amount of time was perhaps slightly different for lower and higher ability students. This makes sense considering that attention spans were probably different for the two groups.

The number of students in each reading group and the number of reading groups per class were also analyzed. The average number of groups seen in a morning showed no relationship to Total Reading achievement, but a positive linear relationship for Word Analysis achievement (4069; $p = .05$, range = 2.62 to 3.75). It may be that a greater number of groups reflect more careful pairing of instruction with ability level. It would be expected that this result also would be reflected in the average group size, because more groups should mean fewer students per group. However, there were no significant relationships for the average group size in the class (4070; range = 5.65 to 8.39).

The number of activities assigned to the students during each reading group was counted. These were worksheet exercises or other work that followed up on material taught during the reading group. The work might be done within the group or taken back to the students' seats to be completed. There were no significant relationships for the average number of activities given during a group lesson (5352; range = .59 to 2.22).

The response opportunity time was coded as to the lesson context used. Five lesson contexts were possible:

1. Slow paced questioning and answers without use of the basal text and without workbooks.

2. Workbook activities in which the students received instructions on performing exercises and/or did these in the group.

3. Fast-paced drill, in which students were expected to respond to the teacher's questions at a very rapid pace, such as in flash card drill.

4. Reading aloud of a new story from the basal text. This involved either silent or oral reading of material which the students had not seen before. It would also include any comprehension questions based on that material.

5. Rereading of a story from the basal reader. This involved either silent or oral reading and/or comprehension questions over material the

students had seen before.

The average time per lesson spent in each of these contexts was computed for each teacher. There were positive linear relationships for the average time spent in Context 1-slow-paced questioning without the reader or workbook (4030; p = .01, .02, range = 2.11 to 9.98). There were no significant relationships for time spent in contexts 2, 3, 4, or 5. Some of these approached significance for one of the two tests; in all such cases, the slope was positive and linear. These results might reflect total time spent, in that greater time overall might be related to more time in any given context.

The proportion of time spent in the different lesson contexts was also examined. The proportion of the total response opportunity time which was spent in Context 1 (questions with no student materials) showed positive linear relationships with achievement (4063; p = .03, .03, range = .10 to .39). There was no significant relationship for the proportion of time spent in contexts 2, 3, 4, or 5. Although not significant, the slopes for Context 2 (workbook) were negative, for 3 (drill) they were positive, for 4 (reading from basal texts) they were negative, and for 5 (rereading stories) they were very close to zero.

These data suggest that the use of the first context is important with its focus on teacher questions with no student materials. These results do not suggest, however, that all time should be spent in that context, because the top of the range represented use of this context no more than 50% of the time. Examining groups of different ability levels or higher grade levels might reveal different patterns of relationships. As students gain ability to read aloud smoothly, they might benefit more from time spent in contexts 4 and 5 (reading of old and new stories from the text) than they would earlier in their reading instruction.

Another measure was time that the teacher spent out of the group once the lesson had begun, and it was expected that this would reflect teachers' management ability. However, there were no significant relationships with achievement for the average time the teacher was out of the group or the average number of times per observation that the teacher left the group (4036, 4037).

Another way of looking at use of time is the rate of response opportunities offered to students. This already has been discussed under Principle 7, but it will be discussed further here. There were positive relationships with gain for the rate of all response opportunities given, and especially the rate of nonturn response opportunities offered. There were no significant relationships with the number of reading turns per minute offered to the students.

The results suggest that the more interactions a child has with the teacher, especially in the form of single questions and answers, the more that child will achieve. This complements the results for lesson contexts, since Context 1, which showed positive relationships with achievement, would allow for more such interactions.

Since there was much emphasis in the model on use of sustaining feedback, it was of interest to know how this affected the overall rate of response opportunities in relationship to achievement. The proportion of response opportunities that were initial selections rather than sustained selections showed no relationships with achievement (645).

Summary of results of time usage variables. These data indicated generally positive relationships with achievement for the amount of time spent in reading group lessons. However, this relationship was curvilinear, suggesting that past a certain point, more instructional time does not yield greater achievement. Other analyses suggest that a format allowing oral teacher questions and answers by individual students is important. Several variables describing such a format showed positive relationships with achievement, while variables describing oral reading or workbook activities in the group were not significantly related to achievement.

There were positive relationships for the number of reading groups seen each morning, although the range was only from 2.60 to 3.75 groups. There were no significant findings for the number of students per group, although this might be expected to vary with the number of groups. These results might represent a careful pairing of instruction with ability level by the more effective teachers. This could result in more groups, but not in equal distribution of students if that was not appropriate instructionally.

Curriculum and Content Covered

Variables included in this section were analyzed at the reading group level, since the curriculum and content covered varied within each class by reading group. Therefore, the $\underline{N}$ for these regression analyses was 66, rather than 20. (Reading groups that were not present for the entire year were not included in the regression analyses, since there were no appropriate test scores for them. Therefore, fewer groups are included here than in the group comparison.) Only Total Reading scores were used as a criterion since earlier analyses at the class level indicated that most results were similar for the two test scores.

191

The content covered during the year showed significant relationships with Total Reading achievement. The reading level completed at the end of the year was positively related to achievement, and there was also a significant interaction (6001; $p < .01$ for interaction and for main effect, range = 1.70 - 3.72, where a score of 4 indicated that the students have completed the first reader in the series, usually considered to be the target for the end of first grade). The interaction effect for this variable occurred because there was a much steeper slope for higher ability reading groups than lower, although the slope was positive for lower ability reading groups also. A comparable variable, the number of basals completed, showed similar results (6006; $p < .01$ for both interaction and main effects, range = 3.45 - 6.10). A reading group that started at the first preprimer and went through all the books considered appropriate for the first grade would have completed six books. Again, although the overall slope is positive, it was steeper for higher ability groups than for lower ability groups. These two results taken together suggest that content covered is definitely related to achievement, and that higher ability students responded best and achieved more at a faster pace than that which seemed most appropriate for lower ability groups. However, even within lower ability groups, those who covered more content achieved more. This interaction is comparable to different results found for high and low SES classrooms at the second- and third-grade level by Brophy and Evertson (Note 1). They suggested that higher SES students (and therefore, presumably higher ability students on the average) benefited from more demands by their teacher, while lower SES students benefited from similar high expectations for performance, but with more emphasis on redundancy and careful coverage of each skill.

There were four basal series adopted by the school district for first-grade reading: Economy, Harcourt-Brace, Houghton-Mifflin and Scott-Foresman. The basal used most in each reading group was noted, and these present-absent scores for each series were compared to the groups' achievement.

There was a significant interaction with entering readiness level for the use of the Economy series (6002; $p = .02$). There was a positive relationship with achievement for higher level groups and a less steep negative relationship for lower level groups.

However, the other three reading series yielded no significant relationships with achievement (6003, 6004, 6005).

There were several measures which examined the use of other materials besides the basal reader as part of the Reading and Language Arts program. There were no significant relationships with achievement for the use of the basal workbook (6007), although since almost everyone used this, the range was restricted. There were also no significant relationships for the use of commercial worksheets (6008), the spelling workbook (6009), the handwriting workbook (6010), or the DISTAR program (6012). There was a weak significant positive relationship with achievement for the use of the English workbook (6011; $p = .08$). Therefore, the use of additional commercial material showed no clear relationship with achievement.

The size of the class and of the reading group was also examined to determine if smaller groups and/or smaller classes were related to achievement. There were no relationships with achievement for the average size of the reading group (6014), but there was a significant positive relationship for average class size for the year (6019; p = .03, range = 20.21 - 28.79). This result suggests that there was greater achievement in the larger classes. This is in contrast to a similar analysis done at the class level, where class size showed no relationship with achievement. It is not immediately clear why the results differ although sample size differences or dependence of sampling units (in the case of the reading group analyses) may be possible explanations. There are analyses currently taking place with other data from this study that may shed some light on problems with different units of analyses (Martin, Anderson, and Veldman, Note 7).

The stability of reading group membership was measured by the number of changes of students during the year and also the relative frequency of change. These demonstrated interactions with the average entering ability of the group. For both variables there was a negative relationship with achievement for more frequent changes within higher ability groups, but no clear relationship within lower ability groups (6015, 6016). These findings suggest that the more stable that higher ability groups are, the better the students will achieve. However, within lower ability groups, there is no indication that stable membership contributes to or detracts from achievement, at least as stability was measured here.

Summary of regression data for curriculum used and content covered. Use of the Economy series (with a phonics emphasis) interacted with the entering ability level of the reading group, so that there were positive relationships with achievement for higher ability groups, but negative relationships for lower ability groups. No more effects were found for use of any basal series. There were no clear relationships with achievement for the use of additional commercial materials. However, there were clear effects demonstrating that more content covered in the reading groups was related to achievement for both low and high ability reading groups, and that a faster pace of content coverage was better for higher ability groups. Reading group size demonstrated no relationship with achievement and class size demonstrated a positive relationship, suggesting that the larger classes achieved more. The variables describing stability of reading group membership suggested that higher ability groups achieve less when there are more changes in reading group memberships, but there are no clear relationships with achievement for lower ability groups.

Other Categories of Academic Teacher-student Interaction

Many categories that describe response opportunities have already been discussed in evaluation of the instructional model and its recommendations for selecting students and providing feedback. However, other categories were also used to describe academic interactions and are discussed below. These were not derived from the instructional model, but we were interested in their relationships with achievement.

Types of questions. Nine categories of questions were included in the coding system. Each type was expressed as a proportion of response opportunities and compared to achievement. There were no significant findings for seven of the nine: repetition, reading choice, word recognition, choice, product, comprehension, or interpretation (666 - 672, 677 - 684).

There were significant relationships with achievement for the use of word attack questions. These were questions which required the student to talk about a sound made by a letter or a part of a word, or to look at a single word and break it into its parts. This variable was examined for total and nonturn response opportunities only, because of low frequency in reading turns. The proportion of total questions that were *word attack* questions showed positive linear relationships (673; p's = .02, .02, range = .04 to .16). This same pattern was repeated for the proportion of nonturn questions that were *word attack* questions (674; p's = .04, .03, range = .07 to .20). This does not suggest that teachers should spend all of their time on word attack questions, because even teachers at the high end of the range only used such questions about 16% of the time. However, teachers who used more of these questions achieved higher reading gains with their students than those who used less.

There were significant negative linear relationships for the proportion of nonturn questions that were *personal* questions (676; p's = .06, .01, range = .00 to .03). This range was restricted, and the distribution was positively skewed so that a few teachers with higher levels of use were probably causing these results. However, the result does suggest that teachers who spent more of the time on personal questions rather than skill-related questions achieved less with their students.

Because there was so much emphasis placed on the use of sustaining feedback in the instructional model, the proportions of each type of question used in sustaining feedback were examined. However, there were no significant relationships with achievement for any of these variables (653 - 659). *Word recognition* questions were used most commonly in sustaining feedback, and they also were the most common type of question asked when all response opportunities were examined.

Types of answers (difficulty level). There were highly significant positive linear relationships for the proportion of total response opportunities which led to correct answers (690; p's < .01, < .01, range = .62 to .77). However, this relationship was not found when nonturn response opportunities were examined separately (691). (Errors in reading turns are examined below.)

The proportion of total response opportunities that had incorrect answers showed a significant curvilinear relationship for Total Reading scores (693; p = .04, range = .18 to .31). These curves are fairly shallow inverted U's for both high and low ability classes. The relationship of this variable with Word Analysis scores was linear and negative (693; p < .01).

The proportion of nonturn response opportunities that included incorrect answers showed significant negative linear relationships with both test scores (694; p's = .08, .02, range = .10 to .16).

There were no significant relationships with achievement for the proportion of response opportunities that led to "don't know" answers for either total or nonturn interactions (696, 697). (These did not happen very often, with the range for total response opportunities being .01 to .02.)

There were significant relationships with achievement for the proportion of response opportunities leading to no response answers for total interactions. This relationship was linear and negative (698; p's = .04, .03, range = .07 to .17). There were no significant findings for the proportion of nonturn response opportunities leading to no response answers (699).

Errors made in oral reading turns. Whenever a group was reading aloud out of the basal text, the observer noted interactions only when a student made an error, and then separately recorded the quality of reading turns as a whole. This means that most of the interactions that occurred during turns came as a result of mistakes, although they included correct answers that followed sustaining feedback questions. The number of response opportunities that did occur per reading turn is an indication of the number of times the reading stopped because of words read incorrectly or not attempted. There were negative linear relationships with achievement for the average number of interactions per reading turn (606; p's = .10, .05, range = 1.33 to 1.89). The number of initial errors made per reading turn reflects the number of original errors made by students and not any errors made during sustaining feedback interaction, and there also were negative linear relationships with achievement for this variable (701; p's = .03, .02, range = .41 to .98).

Correspondingly, there were positive linear relationships for the proportion of all reading turns that were completely correct, without errors (702; p's = .05, .03, range = .53 to .74). There were no significant relationships with achievement for the proportion of interactions occurring during reading turns that included correct answers (703).

These results suggest than when children read aloud, it is better for the material to be easy enough that they can read with few mistakes. However, these results could reflect the fact that students who make more mistakes in reading are more likely to be those who will not achieve as much. It is difficult to say how much these results are due to this factor, and how much they reflect more appropriate matching of material to students by the more effective teachers. (It would be interesting to examine this variable using only a subset of lower level reading groups.) However, because these analyses were done with entering ability used as a covariate, and because interactions were tested for, it cannot be assumed that the results are due entirely to a confounding of ability level with level of error.

Level of errors for different types of questions. In order to examine the level of difficulty in more detail, different types of questions were examined for the proportion of them that led to correct answers. There were significant positive linear relationships with achievement for the proportion of total reading questions that led to correct answers (952; p's < .01, < .01, range = .56 to .75). This relationship was not found when reading questions asked during turn and nonturn interactions were examined separately (953, 954). There also were no significant relationships for the proportion of nonreading questions that led to correct answers (955).

Each separate type of question was examined in the same way. There were significant curvilinear relationships for the proportion of reading choice questions that were correct for both total and nonturn interactions (total: 706; p's < .01, = .01, range = .75 to .89, nonturn: 707; p's = < .01, = .01, range = .75 to .89). In all cases, the curves were very shallow inverted U-shapes. This suggests an optimal moderate point of difficulty for such questions.

There was a positive linear relationship for the proportion of total word recognition questions that were answered correctly (708; p's = .01, < .01, range = .45 to .67). This result suggests that within the range available, teachers who ask easier questions produced higher achievement. This probably is a reflection of optimal matching of material to students.

It does not mean that the teachers never challenged the students with difficult material, but instead, that most of the time (close to 70%), they were asking questions that could be handled by the students. However, this result was not replicated for turn and nonturn interactions analyzed separately (709, 710). It might be that analyzing these variables separately by ability levels would produce significant interactions not apparent with the data aggregated by class.

There was a significant interaction for the proportion of total personal questions answered correctly, as well as for nonturn personal questions. The relationship was positive for higher ability classes, but with a slightly negative or zero slope for lower level classes (714; $\underline{p}$'s = .02, .02, range = .81 to 1.00). These distributions were somewhat negatively skewed, and personal questions did not occur very often, so the results may not be meaningful.

The proportion of product questions answered correctly showed a significant interaction for Word Analysis scores, but no relationship with Total Reading scores (718; $\underline{p}$'s = .04, range = .72 to .84). The relationship with Word Analysis scores was a negative slope for higher ability classes but a zero slope for lower ability classes. This result suggests that higher ability students did better when product questions were not too easy. (These questions could be answered with a single fact or label.)

There were no other significant relationships for the difficulty levels of different types of questions.

Level of errors for different types of selections. Because the instructional model emphasized types of selection and because some types of selections seemed more likely to lead to correct answers (volunteering, call outs), the proportion of each type of selection that did lead to correct answers was examined.

The proportion of ordered selections that led to correct answers showed curvilinear relationships with achievement (640; p's = .01, .08, range = .67 to .84). For both tests, these curves were of a shallow inverted U-shape.

There was a positive linear relationship for the proportion of pre-selections that led to correct answers (641; p's < .01, < .01, range = .60 to .79). There also were significant positive linear relationships with achievement for the proportion of nonvolunteer selections that led to correct answers (642; p's = .03, .04, range = .61 to .76).

Similarly, there were positive linear relationships for the proportion of volunteer selections that led to correct answers for Word Analysis scores (643; p = .05, range = .69 to .84). This result was not significant for Total Reading scores. There were no significant relationships with achievement for the proportion of call out selections that led to correct answers (644). (Most call outs were correct.)

The ranges on these variables, with the exception of call out selections, indicate similar rates of correct answers for each. The general pattern of positive relationships with achievement for higher levels of correct answers was not changed by separately examining the types of selection.

Process feedback. This was coded whenever the teacher gave an explanation about how an answer was figured out. There were significant curvilinear relationships for the proportion of all response opportunities that included process feedback (728; p's = < .01, < .01, range = .01 to .03). All of the curves were inverted U-shaped, although the curves for higher ability classes ascended more steeply before turning. The range for this variable was relatively restricted. It was not examined in turn interactions because it happened so seldom. For nonturn interactions, the same curvilinear pattern was found (729; p's < .01, < .01, range = .01 to .04).

To analyze the use of process feedback in a more detailed way, different types of answers were examined. The proportion of correct answers followed by process feedback again showed significant curvilinear relationships with achievement for total and nonturn interactions (total: 783; p's = .02, .01, range = .00 to .04, nonturn: 784; p's = .02, .02, range = .00 to .04). Again, these curves show the same pattern (an inverted U-shape) with a fairly definite positive slope leading up to a plateau and slight negative slope. Again, the range is restricted, indicating that process feedback did not follow correct answers very often.

Correct answers were broken down into reading and nonreading questions, and there were no significant relationships (866, 867) for reading questions. However, when correct nonreading questions were examined, the proportion followed by process feedback again showed significant curvilinear relationships with achievement (912; p's = .09, .04, range = .00 to .05). Again, the curvilinear pattern is a positive slope that plateaus. The curve for higher level groups rises more steeply before plateauing than that for lower level groups.

The curvilinear patterns did not persist when incorrect answers were examined. Instead, the proportion of incorrect answers followed by process feedback showed significant positive linear relationships for total response opportunities (797; p's = .01, .03, range = .00 to .05), and for nonturn interactions (798; p's = .01, .02, range = .01 to .08). Process feedback occurred more often following incorrect answers than correct answers.

When examined separately for reading questions, the proportion of incorrect answers followed by process feedback showed similar positive linear relationships for total response opportunities (882; _p_'s = .03, .08, range = .00 to .06), and nonturn response opportunities (883; _p_'s = .02, .08, range = .00 to .10).

These results for process feedback suggest that, following correct answers, process feedback is valuable some of the time but should not be overdone. It would be redundant in many cases, and it could seriously disrupt the pace of the lesson, if used too much of the time. However, following incorrect answers, process feedback can be an appropriate technique. The largest range was observed when only incorrect answers to reading questions in nonturn interactions were examined (from 0 to 10% of the time). Perhaps this is the most appropriate time to use process feedback without disrupting the pace of the lesson.

New questions. There was a curvilinear relationship for the proportion of all response opportunities that included new questions (743; _p_'s = .01, .02, range = .10 to .23). The curves are shallow inverted U's, roughly parallel for both low and high groups. When the proportion of nonturn response opportunities that contained new questions was examined, there was a significant positive linear relationship (745; _p_'s < .01, = .01, range = .10 to .19). These results suggest that it is often desirable to follow completed response opportunities with new (typically related) questions, at least within the range observed.

Because the instructional model encouraged use of sustaining feedback rather than terminal feedback, we were interested in the use of new *question* following terminal feedback (so that, in effect, the student's interaction with the teacher was sustained, although not with a question to correct his error). There were no significant relationships found for the proportion of any type of terminal feedback that was followed by a new question (957, 958, 959).

The use of a new question following correct answers was positively related to achievement for total response opportunities (788; *p*'s = .01, .03, range = .13 to .24), and nonturn interactions (790; *p*'s = .01, .02, range - .12 to .23).

The use of a new question following reading turns that were completely correct showed curvilinear relationships with achievement (856; *p*'s = .06, .04, range = .06 to .33). These curves show fairly steep positive slope up to a plateau for higher ability classes, but a shallow inverted U for lower ability classes.

When correct answers were broken down into reading and nonreading questions, there were significant curvilinear relationships for the use of new questions following correct reading questions for total interactions (871; *p*'s = .02, .02, range = .10 to .29), and for nonturn interactions (873; *p*'s = .04, .04, range = .08 to .28). All of these curves showed a positive slope up to a plateau. There were no significant relationships for the proportion of correct nonreading questions followed by new questions (917).

There were no significant findings for the use of new questions following incorrect answers (811, 813, 896, 898, 929). There also were no relationships with achievement for the use of new questions following reading turns that contained some errors (860), or for new questions following no response answers (841, 843).

Summary of results of regression analyses of other response opportunity categories. The types of questions asked, the rate of correct answers, and two other types of feedback not discussed in the instructional model were examined for relationship with achievement. Out of nine categories of questions, only two showed significant results: word attack questions had positive relationships with achievement and personal questions had negative relationships. The difficulty level of questions, as measured by the rates of correct and incorrect answers, did yield significant findings. A higher rate of correct responses, especially during reading turns, was associated with greater achievement. There were curvilinear relationships for some variables describing correct answers, and all variables fell within a range in which some errors were made, so these results suggest an optimal level of difficulty which allows for some errors rather than support for errorless learning. Two other types of feedback, process and new question, yielded positive relationships with achievement, although many of these were curvilinear, suggesting an optimal level. Process feedback was most closely related to achievement when it followed incorrect answers. New questions were most closely related to achievement when they followed correct answers.

Behavioral Contacts

Although not specifically discussed in the instructional model, we were interested in the ways that teachers corrected misbehaviors.

There were negative linear relationships for the proportion of total contacts that were behavioral contacts (1031; p = .01, < .01, range = .09

to .23). Teachers who spent more of their time correcting students for misbehavior produced less achievement, probably because less time was available for academic interactions in their reading groups. Teachers at the high end of the scale had almost one behavioral correction for every three academic contacts.

Behavioral contacts were classified as occurring *within* the group (the teacher correcting students whom she is teaching at that time), or *outside* the group (out-of-group students interrupt the teacher, or she interrupts the group lesson to deal with out-of-group students). Each was analyzed separately as the proportion of all contacts, and revealed similar patterns. Both were negatively related to achievement (in-group behavior corrections: 1077; p = .05, .02, range = .04 to .11, out-of-group contacts: 1078; p = .01, p < .01, range = .04 to .12).

Each behavioral contact was identified as to the type of student behavior involved, and variables were created expressing the proportion of all corrections related to each type of misbehavior. There were no significant relationships for any of these variables, indicating that no one type of misbehavior was more closely related to outcome than any other (1102 - 1012). This was true even when single behavior types were clustered together to express the proportion of behavioral corrections that had to do with interactive and potentially disruptive problems, versus those that were noninteractive and nondisruptive, such as daydreaming (1011, 1012).

Teacher reactions to misbehavior. The teacher reaction expressed in each correction was coded as *management* (a mild, matter of fact statement to the child), *warning* (a more severe correction with evident irritation in the teacher's voice), a *criticism* (a very harsh correction with or without punishment), or *nonverbal intervention* (the teacher corrected the student with an expression, gesture, snap of the fingers, or touch).

There were no significant relationships for the proportion of behavioral contacts that included *management* statements (1015). The range for this variable (.45 to .64) indicated that management was the most common type of teacher correction.

There was, however, a significant negative linear relationship for the proportion of behavioral contacts that included *warnings* (1016; $p < .01$, < .01, range = .12 to .26). This relationship might indicate that teachers who used a lot of warnings did so because they had many behavioral problems and/or because they were overly reactive teachers who responded too harshly to such problems. However, when the absolute rate per minute of *warnings* was examined, there was a significant interaction for the Word Analysis test, with a negative slope for higher classes, and a positive slope for lower classes (428; $p = .02$, range = .01 to .06).

There were no significant relationships for the proportion of behavioral contacts that included *criticism* (1017). The range for this variable was from .09 to .22, indicating that it did occur fairly often, but was not the most common type of correction.

The proportion of contacts involving *nonverbal intervention* showed significant positive relationships with Word Analysis achievement (1018; $p = .04$, range = .01 to .10). However, this distribution was positively skewed.

These results indicate that too much use of *warning* (moderate irritation in correction) is dysfunctional, and some use of nonverbal intervention is appropriate. The ranges observed suggest that teachers are most likely to use management, then warning, then criticism, and then nonverbal intervention. The more effective teachers had more nonverbal intervention and less warning than those who were less effective.

Some separate types of misbehavior were examined for the types of teacher response to them. The only ones that occurred with enough frequency to be analyzed separately were those considered to be individual inappropriate behaviors followed by different types of teacher reactions were similar to the overall patterns just described. That is, there were no significant relationships for use of management, but there was a significant negative relationship for warning (1026; p's = .01, .02, range = .12 to .27). There were no significant relationships for criticism or nonverbal intervention (1027, 1028).

The interruptions involving out-of-group children were examined in some detail to see what effect they might have had on achievement. These were classified as teacher-initiated (she corrected a student in the room) or child-initiated (a student came up to the group from outside and interrupted the teacher). Each of these categories was further divided into brief or long duration.

When only behavioral corrections were considered, the proportion that was in-group rather than out-of-group was not significantly related to achievement (1040). However, when all teacher-initiated contacts were considered (this included all response opportunities, all in-group behavior contacts, and all out-of-group behavior contacts that were initiated by the teacher rather than by the students), there was a significant negative linear relationship with achievement (1079; p's = < .01, < .01, range = .02

208

to .08). That is, considering only those interactions over which the teacher had direct control, the higher the proportion of times that she initiated something with students outside the group, the lower the class achievement. This undoubtedly is a reflection of an overall management ability to have the rest of the class running smoothly while the teacher concentrates on the reading group.

When only child-initiated out-of-group contacts were examined, the proportion that were *brief* rather than *long* was not significantly related to achievement. This also was true for the proportion of teacher-initiated contacts that were *brief*. However, the absolute rate per minute of long teacher-initiated out-of-group contacts showed significant interactions (424; p's = .03, .03, range = .01 to .05). There were negative slopes for higher ability classes, and slightly positive slopes for lower classes.

The proportion of out-of-group contacts that were *child-initiated* rather than *teacher-initiated* was examined to see if one type of interruption was less of a problem than the other. However, there were no significant relationships for this variable (1043).

Teacher reactions to out-of-group behavioral contacts also were examined separately for child-initiated vs. teacher-initiated contacts. These variables reflect the general pattern reported earlier for teacher reactions. Responding to *child-initiated* contacts with a *warning* showed an interaction such that there was a positive slope for lower ability students and a negative slope for higher ability students (1047; p = .07, < .01, range = .00 to .15). This suggests that when students approach the teacher for either brief or long interactions, it may occasionally be appropriate to be severe with lower ability students, but not higher ability students. It may be that this is a type of misbehavior that teachers should discourage, strongly if necessary, especially in lower ability classes.

There were no significant relationships for use of criticism in response to child-initiated contacts, but there was one significant positive linear relationship with Word Analysis for the use of nonverbal intervention (1049; p = .03, range = .00 to .09).

All teacher-initiated contacts were also examined for the proportion that involved each type of teacher reaction. There were no significant relationships for the use of management (1063), but there was a significant linear negative relationship for the use of warning (1064; p = .02, .02, range = .19 to .40). There were no significant relationships for criticism or for nonverbal intervention (1065, 1066).

Summary of regression analyses for behavior contacts. Classes with higher proportions of behavior contacts had lower achievement. Both contacts within the reading group and those occurring outside of the reading group showed negative relationships with gain, especially contacts initiated by the teacher herself to students outside the group. Specific types of student misbehavior were not differentially associated with achievement, but there were significant relationships for the types of teacher reactions to misbehavior. Use of warning statements (moderately severe corrections) showed negative relationships when expressed as the proportion of all behavior contacts. However, there were no significant findings for types of corrections which were less severe (management statements) or more severe (criticism). Use of nonverbal corrections was positively related to gain within a small range of occurrence. There were a few suggestions that more frequent use of more severe corrections (warnings) in lower ability classes might be related to greater gain (within a restricted range of use in the first place), but these were not found for all variables, and were mostly confined to absolute rates rather than proportion variables.

Chapter 5: Summary of Results and Revision of the Instructional Model

The major questions asked in the First-grade Reading Group Study were: 1) Did the treatment have an effect on student achievement? 2) Did the treatment have an effect on teacher behaviors? 3) What were the relationships between teacher behaviors and student achievement?

Chapters 2, 3, and 4 have addressed these questions separately by presenting data for all variables. Based on the results reported, general answers to those questions can be offered, but the final evaluation of the experimental model requires an integration of several sets of results. Therefore, this chapter discusses the groups of variables, proposes modifications in the instructional model based on the results, and identifies patterns of relationships that run throughout the data.

As was discussed in Chapter 2, the adjusted class mean scores on reading achievement were significantly affected by treatment group membership. Therefore, other analyses were performed to determine if this effect could indeed be related to specific parts of the treatment. This required comparing the behaviors of the treatment and control teachers, as well as comparing other aspects of their classrooms and schools. If there were consistent differences in the behaviors between the treatment and control groups in the direction predicted, and if those behaviors were related to achievement, and if no other factors were uncovered which also differentiated the two groups and were related to achievement, then the treatment could be said to have influenced the learning of the students.

Unfortunately, there are no clearcut and unqualified conclusions. It does appear that the treatment was successful in changing teacher behaviors, although not all components of the instructional model were implemented as expected. Many of the behaviors described in the model were indeed related to achievement, and the results of regression analyses shed some light on the contextual factors influencing those relationships. Therefore, there is some support for concluding that the treatment did influence the teachers to behave in ways that were related to achievement.

In general, the major principles underlying the complete instructional model were supported: individual students should have opportunities to receive information from the teacher and practice new skills, receiving feedback on their progress. However, the results suggest some additional ways to achieve this in the small group setting, and some of the specific results lend support to principles not included in the treatment itself, which calls into question how much the treatment actually contributed to the observed group differences in achievement.

Of course, one possible explanation for an overall treatment effect on achievement could be that a Hawthorne effect was present. That is, perhaps the treatment teachers were simply trying harder to teach well, because they knew that they should be doing better since they had received a special treatment. If this occurred, it could account for the relationships of treatment principles to achievement. That is, the treatment teachers could have used the behaviors recommended (such as ordered turns), but also could have tried harder in general, and therefore, achieved more. The specific behaviors (such as ordered turns) would then be related to group differences in achievement, but would not be the causes of these differences. Such a Hawthorne effect may account for other differences between the two observed groups on measures that cannot be related directly to the treatment, such as content covered and level of correct answers.

Recent work by Good and Grouws (Note 8) indicates that gains in achievement may be influenced simply by strong encouragement and knowledge of future evaluation based on observation. Therefore, the possibility of some Hawthorne effect should not be ruled out in evaluating these data.

However, the data describing the treatment group that was not observed suggests that a Hawthorne effect does not completely account for the achievement differences. The unobserved treatment group might be expected to be less subject to such expectancy effects, because they only came in contact with the experimenters at the beginning and end of the year. Therefore, it was expected that they would be less aware of the experimenters' expectations than those teachers who were observed once a week and who knew that the treatment should help them teach better. However, the treatment-unobserved group had significantly higher achievement scores than the control group, but was not significantly different than the treatment-observed group. If a Hawthorne effect was primarily responsible for the differences in achievement, one would expect that the group who remained most aware of their role in the study and our expectations for them (the treatment-observed group) would have the highest achievement.

What, then, accounts for the differences between the groups? On many (although not all) of the variables that were related to achievement, the treatment-observed group did demonstrate different levels of the behaviors than the control group, in the direction related to greater achievement. Therefore, in order to evaluate the effect of the treatment, one must examine the "match" between the group comparison data and the regression analyses. Figure 2 summarizes the key variables in this way.

The following sections examine each component of the instructional

Figure 2:

Summary of major results of analyses of group differences (treatment vs. control) and relationships with achievement

Principle #	Variable	Group differences in expected direction detected:	Relationship with achievement detected:
1	Use of signals in transitions	no	no
1	Efficient transitions	yes	yes (+)
2	Appropriate teacher seating	no	yes (+)
3	Use of overviews	no	yes (+)
4	Presenting new words at beginning	no	no
5	Repeating new words before using them	no	no
6	Sufficient explanations	no	no
7	Providing opportunities for practice to individual students	yes	yes (+)
7	Minimizing choral responses	yes	yes (+)
8	Using ordered turns to select respondents	yes	yes (+)
9	Calling on students for comments	no	no
10	Minimizing volunteers	yes	no
11	Minimizing call outs	yes	yes (+)
12	Avoiding confusing questions	no	no
13&14	Breaking up the group due to different learning rates	no	no
15	Using a student as a model	no	no
17,18,19	Use of sustaining feedback to improve student errors and initial failures to respond	yes	yes (+)
19	Minimize calling on other students to correct an error	yes	no
20	Repeating correct answers	no*	no

Principle #	Variable	Group differences in expected directions detected:	Relationship with achievement detected:
20	Omitting feedback after correct answers	no*	yes (+)
21	Moderate use of praise	yes	yes (+)
21	Specific use of praise	yes	yes (+)
22	Specific use of criticism	no	yes (+)

Variables not directly tied to the instructional model:

	Group differences significant:	
Use of clues when presenting new words	no	yes (+)
Time spent in lessons	no	yes (+)
Time spent in question/answer format	no	yes (+)
Curriculum used	yes	yes (**)
Content covered	yes	yes (+)
Use of word attack questions	no	yes (+)
Success rate of students	yes	yes (+)
Presence of behavior corrections	yes	yes (-)

---

* There were group differences, but in the unexpected direction.

** Interaction

model in these terms, noting when the differences between the two groups corresponded to the regression data in a way that could account for the treatment group's superior achievement. Suggested revisions of the instructional model are then offered. In some cases, such revisions involve describing the desired behavior more completely in order to boost implementation. In other cases, the principle should be changed or eliminated. Other measures not directly related to the original model suggested some important additions.

Getting and Maintaining the Students' Attention (Principles 1 & 2)

The implementation data for these principles showed some differences between the treatment and control groups, but most of these could not be attributed directly to the treatment. Treatment teachers had slightly more efficient transitions, but they were not using signals more often, as was suggested in the instructional model.

The regression data matched the implementation data in that efficient transitions were related to achievement, while measures of the specific suggestions made in the treatment were not. The time spent in transition, student attention to signals, the need for corrections of students once they were in the group, and the time taken by the teacher to get the lesson started were aspects of transitions that were related to achievement, probably because they reflected overall management skills that were reflected in transitions.

Implementation data on seating arrangements as discussed in Principle 2 showed no differences between the treatment and control groups in the way the teacher positioned herself, but a slight difference in the way the students positioned themselves. On the average, however, neither group showed high implementation of this principle. Means for teacher seating indicated that the average teacher could only see 60% of the other students in the room, while 60% of the students could also easily see the rest of the room. Regression data for seating arrangements showed positive relationships for the teacher seating variable, suggesting that the more effective teachers were those who placed themselves to monitor a greater portion of the class. However, there were no significant relationships for ratings of student seating (i.e., how many other students were visible to the children in the group).

Therefore, we cannot conclude that the treatment teachers were more effective due to stronger implementation of the principles related to getting and maintaining student attention at the beginning. However, the findings for efficiency of transitions do suggest that the treatment teachers' greater achievement gains might have been due to better management skills. It may be that the treatment teachers were better classroom managers to begin with, despite the random assignment of schools to treatments. However, it is also possible that the discussion of transitions in the treatment model may have made the treatment teachers more aware of transitions, and this awareness may have resulted in better management of them.

Given the results of the regression analysis, these principles in the model should be revised in this way:

> --Time spent in transition between reading groups should be minimized, and activities should be as efficient as possible. The teacher should concentrate on teaching

the students to respond immediately to her signal, so that corrections are not necessary. The teacher should make sure that she is ready to begin the lesson, once the students are in the group, so that she can capture their attention immediately and spend less time organizing materials and getting the students settled down. When this is achieved, it will result in more time available for content instruction.

--The teacher should arrange the classroom and the reading group area so that she can see as many students as possible during small group instruction.

These two principles should be illustrated with specific suggestions for accomplishing these goals (e.g., how to teach students to move through transitions; how to monitor the rest of the class while teaching the small group).

Introducing the Lesson and New Material to the Students (Principles 3,4,5 & 6)

There were very few differences between the treatment and control groups in implementation of these principles. Treatment teachers were likely to present more new words per lesson and use more phonetic clues when presenting them, but these differences cannot be related to the treatment. (They may be due to differences in the basal series used.) Also, treatment teachers were more likely to dismiss their students to their seats after demonstrating an activity, which was unexpected. Control teachers were more likely to have students begin their written activities while still in the group.

The regression data did suggest some relationships between achievement and behaviors measured for these principles, but these relationships did not usually match the differences between the treatment and control groups. Teachers who failed to give overviews produced less achievement, suggesting that use of overviews to begin a lesson is a good strategy, as was suggested in the treatment model. None of the measures describing the content of overviews yielded information about what kind is best. There were no significant findings for the number of new words presented, although there were treatment effects here. There were positive relationships for achievement with the use of clues when new words were presented, especially for the use of phonetic clues. This relationship matches the implementation data, in that treatment teachers did use more phonetic clues. However, this cannot be attributed to the treatment, since no suggestions were made in the model about using clues here. It might be that focusing the treatment teachers on presentation of new words increased their use of clues, and phonetic clues were probably the most sensible to use when presenting most of the words. The differences between the groups in the basal text used may also have accounted for this, since more treatment teachers were relying on a text that emphasized phonetic rules (the Economy series).

There were no findings in the regression data to support the principle suggesting that students should repeat new words when they are given.

There were also no findings in the regression data regarding the quality of demonstrations and explanations given. It may be that the observers were not sensitive to differences in quality. Also, no distinctions were made between demonstrations of new activities and familiar ones. Perhaps if these had been separated, a relationship would have been detected between achievement and teacher ability to give clear explanations, especially for new or unusual activities.

An interesting interaction with ability was found for one of the variables describing the way in which teachers check student comprehension of explanations. Asking the students to demonstrate the activity to the teacher showed positive relationships for lower ability classes, but curvilinear relationships in higher ability classes, suggesting that some use of this technique is appropriate for all students, but that high levels of use are less appropriate for higher ability students. This makes sense if one assumes that the higher ability students will understand most explanations more quickly, or will be able to question the teacher about areas of misunderstanding. There were no relationships found for other ways of checking student comprehension of explanations, although there were differences between the treatment and control groups on these measures.

Therefore, the data for this set of principles did not suggest that the treatment itself contributed to the achievement differences. Some relationships with achievement were found, but they did not match the implementation data, although one principle (use of overviews) was supported.

Based on the regression data, the principles in this section should be revised in this way:

--Use some kind of overview to begin most lessons. The content of the overviews will vary, depending on the purpose of the lesson and the needs and interests of the students.

--When presenting new words to the students, the teacher should do more than say the word for the student and move on. Much of the time, it will be appropriate to present the word to the student and offer phonetic clues to help the student decode the word, even if the teacher actually gives the word to the child.

--The teacher should occasionally have the students demonstrate to her how they will accomplish the activity before they are allowed to work on it independently. This is especially important with lower ability students and with explanations that are detailed and possibly confusing.

The principles discussed in this section are probably more likely than others to be subject to context effects, and further research done with them should distinguish between the complexity of information given to students in overviews and explanations, indicating whether that information is given when introducing a new skill or in reviewing an old one, and how the student will use the information (i.e., will he be working independently, without access to immediate feedback, or will he be working directly under the teacher's supervision?).

Calling on Individual Students in the Group (Principles 7,8,9,10,11, & 12)

There were two sets of findings that demonstrated both a strong treatment effect and a strong relationship with achievement for this group of principles: rate of individual response opportunities and selection of students. Also, several other variables (not discussed in the treatment model) were related to achievement in ways that strengthened conclusions about these principles.

The first set of findings concerned the rate at which response opportunities were offered to the students. Treatment teachers had higher rates, and there were positive relationships with achievement, especially when nonturn questions were examined separately. That is, the more often the student was allowed to interact orally with the teacher about reading skills through single questions, the more he learned. This finding can be related to the data describing lesson contexts. There were positive relationships with achievement for the use of a context that focused on single questions asked by the teacher and answered by individual students, rather than relying mostly on oral reading or written work in the group. However, there were no differences between the treatment and control classes on these measures. (None were expected, since the treatment did not discuss lesson context.) There were also significant relationships with achievement for the total amount of time spent in the reading group, although there were also no differences between the two groups on this.

All of these results indicate support for greater opportunity to learn and practice skills in oral interaction with the teacher. Thus, these data add to the results of other studies suggesting that achievement in basic skills in the early grades may be optimized by maximizing active student engagement with academic content (Rosenshine and Berliner, 1978). Other data from this study describing content covered also suggest that teachers who wish to optimize reading achievement should set aside much time for instruction and should use that time in as task-oriented a manner as possible.

However, the only result that can be tied directly to the treatment is the finding for rate of response opportunities. Even that was not defined precisely in the treatment, although teachers were encouraged to offer practice opportunities and feedback to individual students.

The results for choral and group responses are related to those for individual response opportunities, and can be directly tied to the treatment. The treatment teachers used fewer choral responses and allowed fewer group call outs, and the regression data suggested that such practices were related to achievement. However, the results also indicated that choral responses and group call outs may mean very different things. The data suggest that choral responses should be minimized, but that occasional call outs by the group, especially in lower ability classes, may serve useful functions.

In other teaching programs (for example, the Direct Instruction model of Becker and Engelmann, in press), choral responses are recommended as a means of teaching in a small group. We do not feel that these results entirely contradict the recommendations of that program, however, Instead the data suggest that the purposes and effects of group responding in various contexts should be examined more closely. It may be that some teachers in this study were using choral responding as a substitute for more active, closely monitored individual practice. However, in the Direct Instruction program, choral responses are used as a method of encouraging active practice, and teachers are trained to be aware of whether or not the students are involved. Therefore, the critical point is not whether group responses are used, but rather how involved and attentive the students are.

One might question whether the higher rate of response opportunities in the treatment group is actually a treatment effect. Principle 7 did not specifically say to the teachers to ask more questions, but it did indicate that it was important for individual students to receive practice and feedback from the teacher, and therefore, choral responding should be minimized. Lowering the rate of choral responding could not account for the higher rate of individual response opportunities, since choral responses did not occur very often, even in the control group. Therefore, it may be that the treatment teachers would have provided more opportunities to individuals to respond, even without the treatment. Again, the curriculum differences must be taken into account. It is possible that the series used by most of the treatment teachers encouraged more individual response opportunities by focusing on decoding skills. Further analyses are being done to determine the relative effects of the treatment and the basal series on such variables as rate of response opportunities and time spent in reading group.

The second important set of variables derived from Principles 7 through 12 which showed consistency between the implementation and regression data were the selection variables. There were very strong treatment effects for the use of ordered selection, in that treatment teachers used this most of the time. Control teachers did not use it often at all, and instead relied on more random selection of students or student self-selection (volunteer or calling out). The regression data suggested a very strong relationship with achievement for ordered turns, a strong negative relationship for the use of unsystematic teacher selection, and negative relationships for students calling out answers. Although we had suspected that strong reliance on volunteering would be negatively related to achievement because the more reticent students would have fewer chances to practice skills, this relationship was not found. However, there were no teachers who relied on volunteers the majority of the time, and it might be that a larger range of behavior would demonstrate some relationship with achievement.

Conclusions from these selection data are supported most strongly for ordered turns, where all variables revealed positive, noninteractive linear relationships with both measures of achievement. However, caution should be observed in interpreting these results. Because there was a very strong treatment effect for the use of ordered turns, it may be that they were highly correlated with something else in the treatment that also was strongly associated with achievement. That is, even though these data come from an experimental study, they are still basically correlational, and cannot be interpreted as meaning that ordered turns caused the higher achievement. However, there are some reasons to believe that ordered selections may be causally related to higher achievement, when other good teaching practices are also present.

First, they equalize the distribution of response opportunities, and insure that everyone gets practice and is therefore exposed to the skills being learned and tested. Second, they help teachers maintain control of the reading group. By reducing the frequency of over-eager students constantly volunteering, and of other students calling out answers, the teacher spends less time trying to decide who to select or correcting students for call outs. Also, the teacher does not have to worry about remembering who has or has not answered questions about a particular skill.

Ordered turns may simply represent the most efficient way for the teacher to remain in control of who answers questions, and yet also insure that all students receive equal attention. The main purposes of the ability-based small group as we see them are to present new information and opportunities to the students to practice and receive feedback on developing skills. Within this particular context, it is not surprising that the advantages offered by systematic selection are reflected in higher achievement.

Revision of the principles in this section would involve deleting some which showed no relationship with achievement and no differences in implementation: those dealing with the use of comments and confusing questions. It might be that the use of comments (asking another student to comment on an answer) could occasionally be a useful technique, and should be suggested to teachers. However, systematic use of this was not supported by the data. Likewise, although it seems sensible that confusing questions should be avoided, the data indicated that this is not a large problem, and except for individual cases, should not require special emphasis in a treatment.

The suggestion given in Principle 7 about providing feedback to every answer should be modified to suggest that it is important to offer feedback to all answers that are not correct, and probably also to correct answers that are not obvious to the students. However, teachers seem able to judge when correct answers do not need feedback and are probably showing good judgment about pacing when they omit it occasionally. This may be due to sensitivity to pacing, or it may indicate something about the types of questions that were asked by the more effective teachers, who were more likely to omit feedback to correct answers. It seems futile to simply measure the presence or absence of feedback without examining the information needs at that point in time. Probably a better way to approach this would be to define an underlying principle that students always need to know when an answer is correct and sometimes need to know what was correct about it. The teacher should remain aware of the students' needs for information about their answers, and provide it when necessary.

A revised version of the model would include similar suggestions regarding selection of students to respond. One amendment would be to recognize that teachers are not likely to correct a call out and also accept it. This suggestion was made in Principle 11, but was not implemented. If call outs were accepted for their content, they were not corrected, and vice versa. However, the treatment teachers did have fewer call outs and the regression data suggest that this was desirable. Therefore, there should be more emphasis in the model on preventing call outs and on what they might represent (in terms of student enthusiasm, etc.) rather than on ways of dealing with them when they occur. Again, there should probably be a discussion of the meaning of call outs in different contexts, to make teachers aware that they may sometimes represent something appropriate, but that at other times they represent control problems. Perhaps a useful modification of the model would be to suggest that teachers develop a clear signal system to indicate to students when they are free to shout out answers and when they are not. Further guidelines could be provided to teachers about ways to evaluate the effects of call outs in order to make decisions about when to use them as a technique to involve the students.

Therefore, the revised version of this section of the model would be as follows:

> --Maximize your students' opportunities to interact orally with you about reading skills through questions and answers.
>
> --Your questions should be directed to individual students almost all of the time. When group responses (choral responses) seem appropriate, make sure that all of the stu-

dents are involved and attentive.

--Always provide feedback to answers that are not correct. Feedback ordinarily should be provided when answers are correct, as well, except when it is obvious that students know they have answered correctly. Here, it will sometimes be appropriate to omit feedback.

--Select students to answer questions systematically, such as going in order around the group. This insures that all students have an opportunity to practice important skills.

--Minimize call outs from students so that you can distribute response opportunities as necessary. Sometimes, however, you may want to allow call outs to encourage interest or pick up the pace. A signal can be devised to indicate to the students when you will allow call outs. At other times, however, they should be discouraged.

Dealing with Individual Learning Rates within the Group (Principles 13,14, 15, & 16)

This group of principles had poor implementation by the treatment group, and the regression data did not show relationships with achievement. They suggested behaviors that were probably unfamiliar to most of the teachers, and it seems likely that the brief explanation given in the treatment material was insufficient to convince the teachers to try them. Another possibility, of course, is that they were not appropriate for first-grade reading groups. Indeed, this set of suggestions was derived primarily from materials produced for teaching in a bilingual kindergarten setting where language learning is more important than reading per se.

In revising the instructional model for first-grade reading groups, these principles would not be included, at least not in their present form. However, work done with them in other settings (i.e., the bilingual kindergarten) indicated that they may be useful in some situations where there is more of an emphasis on language learning.

Responding to Answers that are not Correct (Principles 17, 18, & 19)

Principles in this section yielded many variables in order to tap effects of type of question, relative effectiveness of feedback, and student ability level. As might be expected when such a determined effort is made to complicate matters, the results, although generally consistent, are complex.

These principles suggested that the teacher should provide feedback to incorrect answers (or failures to respond), thus discouraging call outs from other students and not asking other students for the answer, and that whenever possible, this feedback should be sustaining, so that the original student had opportunities to improve his own answer by being asked simpler questions. Implementation data for these principles demonstrated that, as predicted, treatment teachers had a higher rate of use of sustaining feedback, especially the category of clues. They asked other students for the answers fewer times, and they had fewer instances of other students calling out feedback. The two groups were very similar in having the teacher give the correct answer to the student. The treatment teachers were slightly more effective with their sustaining feedback, in terms of it leading to a correct interaction with the students, although this was not emphasized in the treatment itself. It may be that the treatment teachers expected more from sustaining feedback (because of the rationale given in the treatment and because their attention was focused on it) and therefore were working harder to see that it was effective.

The regression data generally matched the implementation data, and did suggest several ways in which the effects of different kinds of feedback are moderated by context. Generally, the higher the proportion of sustaining feedback to terminal feedback, the greater the achievement. However, all teachers, including the most effective, used both sustaining and terminal feedback. Analyses of more precise variables suggested that appropriate use of feedback cannot be defined just by these two categories, but also must include whether it meets the students' need for information about an answer, and whether it interrupts the pace of the lesson. These results are compatible with the treatment, in that teachers were encouraged to use their judgment about when sustaining feedback was appropriate to a question and when it would fit the pace of the lesson. The data did suggest, however, that improvement of incorrect answers or failures to respond through sustaining feedback is a desirable thing, as was emphasized in the treatment.

There were many interactions between the feedback variables and the mean entering ability level of the classes. For example, an interaction was found for call outs from other students, such that there were positive relationships with achievement for lower ability classes, but negative relationships for higher ability classes. This behavior did not occur very often, so these results did not indicate that high levels of called out feedback were related to achievement. The interaction is similar to one found in other research (Brophy and Evertson, Note 1) and in the data on group call outs, and is interpreted similarly. That is, call outs may mean different things in the two different types of classes. Within lower ability classes, they may represent students' enthusiasm and eagerness, whereas in higher ability classes, where lack of motivation is not as likely to be a problem, they may represent a control problem.

Other types of feedback yielded significant main effects. The teacher's giving the answer was related to lower achievement, especially when it followed failures to respond. (However, all teachers did this some of the time.) Two types of sustaining feedback, clue and give by clue, were related to achievement. These results support the general conclusion that sustaining feedback is advantageous much of the time. Teachers need to help students process the information that will help them correct an error or see the reason for it, and giving the student the answer and going on to another question much of the time is not going to accomplish this.

We expected to find some interactions with entering ability for these variables, such that higher ability students would need less of the detailed information offered by clueing feedback and could more often be helped by the teacher giving the answer to them. However, these interactions were not found, and the negative relationship for giving the answer was similarly strong for both ability levels. The only interaction for clueing feedback was found when examining mistakes made during oral reading turns. Here, in higher ability classes, it seemed better for the teacher to simply prod the student or wait for him to correct his mistake, rather than providing clue feedback most of the time. However, in lower ability classes this pattern was reversed. This was interpreted as indicating that lower ability students at this grade level do, indeed, need more information to correct an answer than is provided by simply repeating the question, whereas higher ability students, given more time, may work through to the solution more often.

However, this result might also indicate that the purpose of oral reading turns within higher and lower ability classes is different. If the students in these classes are reading at different levels, and if the higher ability students were reading longer passages and focusing more on comprehension, then the optimal feedback might be determined by what would be necessary to keep the pace going. In this case, stopping to give clue feedback would disrupt the pace, although it would help the student work through the process of sounding out a word. Waiting a few seconds for the student to figure it out on his own might actually be a more efficient way of getting the word read.

All teachers showed a tendency to give the answer to the student in reading turns more often than in other situations, and this probably also reflects pacing requirements, in that giving the answer allowed the student to move on without disrupting comprehension of the passage. When examined only in reading turn interactions, giving the answer to the student was not related to achievement at all, so that the negative relationship found overall for its use may not necessarily apply within reading turns where the pacing requirements are so different.

In the original treatment manual, there was much discussion of how the teacher must use her own judgment in selecting feedback strategies. Distinctions were drawn between the types of questions being asked, in that some answers were obviously not amenable to sustaining feedback.

Distinctions were also drawn between different pacing speeds, in that in rapid drills the teacher would be more likely to give an answer and move on quickly. In revising these principles, the general rationale would remain the same, with teachers being encouraged to use sustaining feedback as much of the time as seemed appropriate, especially when students failed to respond. The discussion of appropriateness of feedback should be expanded slightly to consider the pacing requirements of reading turns as well as drills, since reading turns were more common than drills. There should also be a discussion of how students of different ability levels will vary in their dependency on the teacher for information about their mistakes. In the original model, the description of the behaviors involved were generally adequate, since the implementation for this group of principles was good. Therefore, they need less substantive revision than many others in the treatment model. The only major change would be to eliminate the discussion of asking other students for the answer. This was discouraged, but the regression analyses did not show that it was undesirable. It should be discussed as one type of feedback that should be evaluated in the same terms as the others. That is, does the student receive the necessary information without undesirable effects (e.g., disruption of pace, student embarrassment)?

The results support including the following in a revised model:

> --When students answer incorrectly or fail to respond, the teacher's feedback is very important. Many options are open to the teacher, and all will be appropriate at various times. The teacher must make a decision about how to respond to the students, based on two considera-

tions: a) What information does the student need, and how will he use it? b) How will the feedback affect the overall pacing of the lesson, and how does that relate to the lesson's objectives?

The following principles can be used to answer these two questions:

--Most of the time, it is better to provide sustaining feedback to the student by asking simpler questions that lead him through the process in order to correct the answer himself, or by simply allowing him more time to correct without help. When using sustaining feedback, the teacher should tailor it to the student by making it possible for him to answer the new question with the information provided.

--Sometimes sustaining feedback will not be appropriate, due to the type of question asked (it cannot be broken down into simpler questions) or the pace desired by the teacher (i.e., when students are reading a passage aloud for comprehension purposes or when conducting a fast-paced drill, sustaining feedback would interrupt the pace necessary for the purpose of the lesson). When this is the case, the teacher should provide the correct answer in some way, although she should be careful that students do not come to rely on this, especially if students frequently fail to respond.

Responding to Correct Answers (Principle 20)

This principle emphasized that correct answers should be acknowledged and the teacher should make sure that all students heard and understood them. The instructional model suggested that teachers would occasionally need to repeat an answer, or have it repeated by the students, and would generally want to provide some kind of feedback following correct answers.

Neither the implementation nor the regression data met expectations. The treatment teachers were actually less likely to use emphasis feedback (repeating or having an answer repeated) after correct answers, and they were more likely to omit feedback after correct answers. The results suggest that this principle may have focused the treatment teachers' attention on their behavior following a correct answer, but their own judgment suggested a change in their behavior opposite to what was predicted.

The regression data indicated that the treatment teachers were using good judgment. Although there were no relationships found for the use of emphasis feedback following correct answers, there were positive relationships with achievement for omission of feedback after correct answers. This occurred an average of less than 15% of the time, and no teachers were omitting feedback most of the time. Although it cannot be determined from the data, it may be that the correct answers that were not acknowledged were apparent to the students, so that they did not need the information provided by teacher feedback. The teacher may have been omitting feedback in these cases because stopping to provide feedback might have unnecessarily interrupted the pace of the lesson. However, it is possible that omission of feedback following correct answers is as much a reflection of the type of questions asked, as of the teacher's feedback style. That is, perhaps the more effective teachers (who tended to omit feedback more often) were also asking simpler questions with obviously correct answers.

In revising this principle, the discussion should include the basic premise that all studetns in the group should have heard and understood the answer, but there should be less emphasis placed on repeating the answer. Perhaps it is more important that the teacher maintain the students' attention on the lesson, so that repeating the answer is unnecessary. However, there was no evidence to indicate that repeating the answer is an undesirable thing to do. Therefore, it could be discussed with teachers as a useful technique to use some of the time. The principle should therefore be rewritten as follows:

> --Maintain the students' attention on individual responses so that they will hear correct answers when given. If the teacher feels that other students did not recognize the answers as correct, she should do something to focus their attention on it.

The discussion of this principle of providing feedback to all correct answers should be less absolute. The focus should be more on times when it is appropriate and desirable to omit feedback following correct answers. As is the case with many of the principles, appropriate use is dependent to a great extent on the teacher knowing what information the students need in order to understand the skills being taught.

Praise and Criticism (Principles 21 & 22)

These principles suggested that praise be used moderately and with discrimination. It was suggested in the treatment that both praise and criticism should be very specific, so that students could know exactly what it was about their behavior or answers that was desirable or undesirable.

Implementation data for these principles showed that there was some treatment effect, but it was not strong. Treatment teachers did use less praise and were slightly more specific in the use of praise, but they were not specific very often. There were no differences between the two groups in the level of criticism offered (although Principle 22 did not actually suggest that there be more or less criticism, only that it be more specific). There were no differences between the two groups for specificity of criticism.

The regression data match the implementation data for praise, in that there were generally negative relationships with achievement for praise and one curvilinear finding, indicating that a moderate amount of praise was most effective. However, these results were not consistently strong, and were found for only one of the test scores. There was a positive relationship with achievement found for specificity of academic praise, but this also was limited to one test score.

Praise was not specific very often. Indeed, less than 8% of all interactions that included praise were specific. Praise itself, however, was more frequent. When nonturn interactions were examined separately, the range that included praise was from 4% to 19%. Some teachers were praising one out of every five answers given by the students. For this variable (i.e., nonturn interactions that included praise), relationships with both test scores suggested that those teachers at the high end of the scale (those offering praise almost 20% of the time) were achieving less. Although it cannot be determined directly from the data, it makes sense that such frequent use of praise may have seemed insincere and meaningless. Other research has suggested that frequent praise may indeed represent very low expectations by the teacher, which are also reflected in less effort to encourage good performance (Brophy and Good, 1974; Kleinfeld, 1975; Weinstein, 1976). That is, the teacher praises work that is less than the student's best, because she did not expect any better work. If such an attitude is related to frequent use of praise, it is not surprising that there are negative relationships with achievement.

There were no relationships with achievement for academic criticism. It was rare, occurring in less than 10% of all academic interactions. However, when both academic and behavioral criticism were examined together, there were significant negative relationships with achievement. This result probably reflects the teachers who were overreactive and focusing too much attention on student misbehaviors, probably because they had poor managerial skills. When total criticism (both academic and behavioral) was examined for specificity, there were positive relationships with achievement, although as with praise, there was little specificity (less than 10% of all criticism).

In revising Principles 21 and 22, we would retain the emphasis on a moderate and discriminating use of praise, with suggestions to be as specific as possible with its use. The low levels of implementation of specific praise, however, suggest that the treatment itself should be more precise, expanding this discussion to include more examples, as was done with the section on feedback.

The discussion of criticism in the model should probably be expanded to include more emphasis on preventive management, and to help teachers focus on the meaning of extremely negative cycles of critical reactions to students. Likewise, the discussion of specificity (i.e., pointing out to the students what they should be doing instead of the misbehavior or incorrect answer) should be expanded to provide more examples.

Therefore, the revised principles would read:

> --Praise should be used in moderation and should be as specific as possible about what is being praised. The teacher should be careful not to overdo nonspecific praise, especially if she finds herself responding automatically with praise instead of responding to the answer with more informative feedback.
>
> --Criticism may be appropriate sometimes, although the teacher should be as specific as possible about what desired alternatives are. Criticism of students' behaviors may sometimes result in failure cycles in which the criticism is not effective. Teachers should concentrate more on preventing misbehaviors (by supplying enough work to do easily, by consistently enforcing expectations in clear informative ways, etc.), and remain aware of the effects of criticism when they do offer it. If it does not have the desired effect (changing the students' behavior), then the teacher should seek other solutions.

Time Usage

Several variables were analyzed to see how teachers used their reading group time. Only one of these (rate of questions) was related to the treatment, so few differences were expected between the treatment and control groups. Some differences were apparent, however, but most did not relate to achievement.

The treatment teachers did not differ from the control teachers in the amount of time spent in the small group lesson, or in the way that time was used in different contexts. However, these measures were related to achievement. The more time spent in the lesson, and especially the more time spent in a format designed to present questions to individual students to answer orally, the more the class as a whole achieved. This may be because there was more public demonstration of skills. Even though each student did not answer each question, he heard it.

The treatment teachers did differ from the control group in one way that was reflected in the achievement data: more reading groups were seen each morning (although there were no differences in the average time spent with each group). This might mean that the treatment teachers were more closely matching their instruction to ability group level, but further analyses are necessary to determine this. That is, more homogeneous groups (which would result in more groups in most classrooms) may be better. However, until such analyses are done, this finding should not be the basis for a principle. Data discussed in the next section demonstrated that large classes had higher achievement. It may be that the larger classes also had the most groups, and that the number of groups reflects this, rather than careful matching of students by ability level. Indeed, there were no relationships with achievement for average group size.

There also were no relationships with achievement for the number of activities assigned to students during the reading group lesson. This may not accurately reflect the amount of follow-up work given to the students to do, since the teacher may have made other assignments before the students were seen in the group. Therefore, this variable may simply reflect teacher style in giving assignments, rather than the amount of seatwork assigned to the students. Treatment teachers did assign more activities in the group than did control teachers, but this difference was not reflected in the achievement data.

Therefore, the time usage data suggest some principles that should be added to the instructional model, but they do not indicate that the treatment group and control group differed in ways that accounted for the treatment effect on achievement.

The following principles should be added to the model, based on these data:

> --Reading groups should be between 25 and 30 minutes long. The length will depend on the attention level of the student, which will vary by ability level and probably also by the time of the year and the context of the lesson.
>
> --Up to one half of the lesson time should be spent in a format that allows for teacher questions to individual students, without the use of materials. Oral reading and work in workbooks will sometimes be appropriate, but should not be used to the exclusion of the questioning format.

Curriculum Used and Content Covered

There were differences between the control and treatment groups on the basal series used most often and in use of other commercial materials. There were also differences in the rate at which students moved through the material, with the treatment reading groups moving at a slightly faster pace. Some of these differences were also reflected in the analysis with achievement scores, suggesting that part of the treatment effect may have been due to curriculum and other factors not directly related to the treatment.

Although there were large differences between the treatment and control groups on the choice of principal basal, there were no main effects to suggest that one series was uniformly better than another. However, there was a significant interaction suggesting that one series, Economy, was positively related to achievement for reading groups that started with higher levels of readiness, but was negatively related for lower ability groups. Since more treatment than control reading groups used the Economy series, it is possible that at least part of the overall treatment effect (i.e., greater adjusted achievement in treatment classrooms) was due to greater use of the Economy series. However, the absence of a main effect, and the slightly negative relationship with achievement in lower ability reading groups prevents clear conclusions about the effects of the confounding of curriculum and treatment. However, the Economy series did place greater emphasis on phonetic rules than the other series, and other research (Chall, 1967) has indicated that phonics are important in beginning reading instruction.

It is easier to interpret the data describing the amount of content covered. The teachers who exposed their reading groups to more of the basal curriculum had higher adjusted achievement scores. The slopes for higher and lower ability groups were significantly different, although overall there was a positive relationship. These results suggest that it is important for the teacher to encourage the students to cover as much material as possible, and that a faster pace may be relatively important for higher ability groups. This is comparable to other research done with elementary students (Brophy and Evertson, 1976; Good, Grouws, and Beckerman, 1978; Brophy and Evertson, Note 1).

The treatment did not discuss how fast content should be covered. However, there were weakly significant differences between the treatment and control reading groups on these measures. Since these analyses were conducted with an N of 66, the use of the .10 level of significance is not as justified as it was with the analyses using an N of 20. However, the trend toward a higher level of significance should be noted. It is possible that the treatment teachers pushed their students somewhat faster because they were aware that their performance was being monitored and because they were expected to outperform the control group (i.e., there may have been a Hawthorne effect that was not directly related to the content of the treatment). It might also be that some components of the treatment made it easier for the teachers to move somewhat faster through the books. However, in interpreting the overall treatment effect (i.e., the treatment groups achieving more than the control groups) the data on content coverage must be considered.

In revising the instructional model, one additional component should be an emphasis on presenting materials to the students to maximize their exposure to the skills. The data on time usage suggest that one way to do this would be to extend the reading group time. The suggestions made in the original treatment model, at least those that showed relationships with achievement, are more likely to be useful when they occur in a setting that makes it possible to present more to the students and insure that they are attending to the information.

Therefore, the following principle should be added:

> --Students should be encouraged to move quickly through the curriculum if teachers desire to maximize achievement scores. The rate at which they are exposed to the curriculum will vary, depending on the ability level of the students. However, rapid coverage of the curriculum must include learning of the skills, not page coverage per se.

Other Response Opportunity Categories

In addition to the parts of response opportunities that were most emphasized in the treatment (selection and feedback), the coding system included descriptions of types of questions, level of answers, and some additional categories of feedback.

Types of questions. There were no differences between the treatment and control groups on the types of questions asked. Both groups asked mostly word recognition question about 55% of the time, with the second most frequent type of question being comprehension questions (about 12% of all questions on the average).

There were some relationships with achievement for two categories of questions that did not occur frequently. Word attack questions (the students was asked about a letter or sound within a word) were positively related to achievement. Teachers who asked more word attack questions used them about 15% of the time, compared to an average of 10% for all teachers. It may be that use of such questions was also related to the curriculum used, since one series (Economy) had a stronger emphasis on phonetics. However, use of that series showed an interaction with initial ability to affect achievement, rather than the linear effect shown for word attack questions. Also, there were no differences between the two groups on frequency of word attack questions, even though there were differences in use of that basal series. Therefore, at this point, question type cannot be attributed directly to the choice of basal. (Further analyses may detect such an influence.)

There were negative relationships for personal questions, although these did not occur often (the mean was 2%), and the result was probably due to a few teachers with outlying scores. It seems unlikely that using personal questions such a small percentage of the time would relate to achievement, although the results do suggest that this should not be overdone.

Since there were no differences between the control and treatment groups on the question variables, the treatment effect cannot be related to these two findings. However, revisions of the model should probably include a discussion of word attack questions, in that they may help the student understand better how to apply phonetic rules when reading a word. Earlier reading research (Chall, 1967) has suggested that phonics instruction should be included in early reading instruction, and this type of question is likely to be a reflection of such an emphasis.

Level of answer. There were differences between the two groups for relative frequency of correct answers, and these differences were reflected in the regression data. Therefore, the overall treatment effect may have been partially due to this factor, although this cannot be related directly to the treatment, except for no response answers. The treatment group had a lower rate of failures to respond, and this can be related to the emphasis in the treatment on eliciting a response to every question. Likewise, the regression data supported this part of the treatment: fewer failures to respond were related to greater achievement.

Higher levels of correct answers were related to greater achievement, although within a range in which some errors occurred. That is, there was no support for errorless learning. These results were clearest when reading turns were examined separately, indicating that oral reading should be based on material that is easier for the student than questions asked out of turn. This makes sense, if the purpose of oral reading is to practice skills other than word calling, such as reading with expression and for comprehension. If the level of the material is too difficult, then more attention must be paid to correcting single words at the expense of the comprehension process. However, in a different setting, such as a question-answer exchange with the teacher, errors are apparently not a hindrance to the purpose of the lesson, and indeed may be important for diagnosis about the skills involved.

These results may be taken together with those describing appropriate feedback to errors in order to define two contexts in which errors have different meanings and require different responses from the teacher, both in planning and in conducting the lesson. Oral reading may be viewed more broadly as an exercise in applying several separate skills for a larger purpose, such as comprehension, and to build fluency and speed while doing so. That is, the purpose is one of coordinating several separate skills that need to be almost automatic for the coordination to take place. It makes sense, then, for this to occur at a level where the separate skills (in this case, word calling) are not so difficult for the child that he cannot coordinate them. It also makes sense for the teacher to provide feedback to errors in a manner that does not interfere with the larger purpose of the exercise. The feedback data did not suggest an optimal response to errors in turns as strongly as it did for nonturn response opportunities. Perhaps in reading turns (or any situation where the purpose is to coordinate the application of several skills) the type of feedback is less important than the initial control of the error level (i.e., the teacher should make sure that the separate skills can be used easily in that exercise).

This situation is very different from one where the focus is on learning separate skills (such as applying a single phonetic rule to decode a word). In the First-grade Study, nonturn response opportunities are examples of this: for nonturn response opportunities, there were no significant findings for error rate, while there were more important relationships for the type of feedback following errors. Perhaps, then, in lessons in which the purpose is to build skills before having to coordinate them, the error rate is less important than what is done with the errors (i.e., the type of feedback offered by the teacher) so long as it is not too high.

There are no apparent reasons why the treatment group had higher rates of correct answers than the control group, since this cannot be directly related to the treatment. As was discussed in the section on content coverage, the treatment teachers also moved their students through the basal materials slightly faster, and therefore can be described as more challenging, even though the success rate was higher. They were apparently better at designing instruction that balanced the right amounts of challenge and success that would allow for easy progress. This might have been due to pretreatment differences, or it might have been due to a Hawthorne effect in that the teachers were "inspired" to be better teachers. The general tone of the treatment materials could have contributed to this, but the specific behaviors in the model are not directly relevant to the findings for error rate and challenge. In any case, a revised version of the model should include a discussion of error rate and how it is important.

Other types of feedback. The two additional types of feedback that were examined yielded significant relationships with achievement, although there were no group differences. Process feedback, in the form of an explanation about how to get the answer, was an appropriate thing to do, although it was not used extensively. It seemed especially useful after incorrect answers. This can be related indirectly to the treatment, in that the discussion of sustaining feedback in the instructional model emphasized the students' understanding of the process necessary to correct an answer.

Therefore, additional information should be included in the part of the model describing feedback to incorrect answers about offering explanations occasionally when the students have not applied the process correctly.

Also, placing new questions to the same student who correctly answered the previous one showed positive relationships with achievement, especially after correct answers. Although they were not coded as such, most new questions were related to the previous questions, so that they represent a kind of sustaining feedback, although not for the purpose of correcting an error. Instead, they may serve the purpose of extending a line of thought through a series of questions to the same child. There were curvilinear findings for some of the variables describing new questions, which indicate that it should not be overdone, probably because staying too long with one student may mean losing the attention of the others.

Therefore, revisions of the model based on the results of these other categories of response opportunity variables would include:

--Include some questions that focus on word attack skills, up to about 15% of the time. Be careful not to overdo questions about personal experiences at the expense of skill-related questions.

--Be attentive to the number of errors that students are making and how that affects the purpose of the particular lesson. If students are practicing skills that they should know well, but are working on using them rapidly and smoothly, then the material should be relatively simple. When students are working on new skills and receiving frequent feedback from the teacher, the difficulty level is not as important, although most questions should elicit correct answers. When errors do occur, apply the principles describing feedback to incorrect answers.

--In general, feedback to any type of answer that emphasizes the process to solution or the steps in a sequence will be appropriate much of the time. The principles describing sustaining feedback to errors are based on this premise. Additional feedback techniques that are similar in purpose and which may be used on occasion are process feedback, in which the teacher explains to the students how to get an answer, and asking a new question to the same student, to extend a line of questioning to its logical conclusion.

Behavior Contacts

This group of variables showed some differences between the treatment and control groups, and also some relationships with achievement that match those differences, although the correspondence is not perfect.

Treatment teachers had slightly fewer corrections than did control teachers, although the results were marginally significant in many cases. Differences were especially apparent for interruptions to correct a student out of the group: control teachers did this more. It is possible to relate this to the treatment, since Principle 2 emphasized sitting to monitor the rest of the class, although general management of the entire classroom was not emphasized as such in the treatment model. Here again, there are differences between the two groups that are open to several interpretations: there might be a treatment effect, there might be a Hawthorne effect, or the treatment teachers might have been better managers to start with.

The frequency of behavior contacts, both within the group and to students in the classroom, was negatively related to achievement, supporting the contention that classes in which behavior problems are prevented will achieve more. (Generally, the absence of behavior contacts indicates the absence of problems, although it is of course possible for a teacher to ignore misbehavior, in which case no behavior contacts would have been recorded.)

The type of misbehavior was not related in any way to achievement, even when examined as cluster variables of disruptive vs. nondisruptive behavior. Most behaviors leading to correction within the group were nondisruptive (i.e., they involved single students being inattentive, misusing materials, etc.).

The results for teacher reactions to misbehavior did not reveal differences between the treatment and control groups, although there were some relations with achievement. Teacher reactions could be rated on a scale of severity, from nonverbal intervention, to a management statement, to a warning statement, to criticism. It might be expected that more severe statements would be negatively related to achievement, since they might represent management problems. This was true for warnings, where negative relationships were found. However, there were no significant findings for criticism alone, although this did not occur very often. Teachers who responded to misbehaviors most of the time with irritation (which would have been coded as warning) were probably those with the most problems in managing the classroom and/or who overreacted to students, which could dilute their credibility. However, these differences in teacher reactions were not reflected in the group differences as were the proportions of misbehaviors. (Since the treatment did not address this topic, except to recommend that criticism be specific, no differences were expected.)

There were not extensive data collected on overall classroom management since this was not emphasized in the model. However, these data suggest that it is important, since the teacher must be able to concentrate on instructing the students in the small group, and this involves maintaining their attention as well as preventing interruptions from out of the group. Also, it is likely that overall reading achievement depends not only on the group instruction, but also on the independent work done while other groups are being taught. Therefore, the teacher who establishes an atmosphere in which more can be accomplished by students working on their own while she is with a group is likely to maximize their practice of skills (and hence, learning) in two different ways: through engagement in seatwork and effective instruction in the group.

The revision of a principle describing use of criticism has already been discussed (see page 230) and is further supported by these data. Another important principle can be derived from the data describing behavioral contacts.

> --Stay aware of how many of your contacts are behavioral in nature, and work to prevent misbehaviors so that more of your attention may be focused on the lesson. There are two places where this is important: with the students in the group, and with students at their seats. Minimize interruptions of the group lesson to deal with out-of-group matters.

Unfortunately, this is easier said than done, and to be effective such a principle would have to be illustrated with specific suggestions for improving management. Although the First-grade Study did not yield such suggestions, the work of Kounin (1970) is a valuable source, and other research presently being conducted by the authors may provide relevant information here (Evertson and Anderson, Note 9; Anderson and Evertson, Note 10).

Discussion

These results suggest a useful teaching style for small group instruction in the early elementary grades. Many points in the instructional model were substantiated, and many of the other variables yielded additional information. Taken together, the data presented a fairly cohesive picture of the "effective first-grade reading group teacher." Some of the principles embodied in the results may be applicable in other primary-level classes for other subjects and other formats, but many of the specific techniques that exemplify the principles may be less appropriate in a different context (e.g., large group instruction, math instruction, third graders). For example, the principle of providing many practice opportunities to students who are learning basic tool skills is probably important in many settings, but the use of ordered turns to insure systematic selection of students may be less appropriate with large groups or with a lesson where the content is more predictable, especially with older students who might indeed "tune out" until their turn.

Even though the appropriateness of specific techniques will change with the context, the value of the general principles is not negated. Indeed, a useful approach to organizing research to use in teacher education may be to identify the important major principles, such as those discussed below, and then to describe for teachers the specific techniques and strategies that will embody those principles in different settings. The instructional model used in this study was an attempt to do just that: to begin with the principle that the teacher must manage the entire group while at the same time (and, in fact, in order to) provide attention to individuals within the group. Another underlying principle related to this was that the teacher must elicit responses from students to allow them to practice skills being taught and receive immediate feedback before being allowed to work on their own. The specific suggestions for accomplishing this were designed to be appropriate to first-grade reading groups, and many of the results of the study can be used to make them even more appropriate in that setting. (For example, the original model did not emphasize different pacing requirements of reading turns as opposed to questions asked outside of turns, and yet the results suggest that this is an important factor to consider in defining appropriate error rate and feedback.)

Patterns in the data suggest that the following principles are valuable and that their implementation (in a manner appropriate to the specific setting) will foster student achievement:

1) As other research has found (Rosenshine and Berliner, 1978; Good, et al., 1978), students achieve more when they were given greater opportunity to learn. In this study, the more effective teachers spent more time with the group and covered more content as a result. However, no teacher met with reading groups an inordinately long time; for first graders that would have been inappropriate. However, the teachers whose students achieved more met with their groups for 30 minutes as opposed to 20 minutes a day. Although measures of seatwork were not taken in this study, it is likely that the careful matching of follow-up work is also important, with sufficient time being spent by the student to practice skills to the point of fluency. There were indications that classes in which students at their seats worked quietly and independently also achieved more, which does suggest that time-on-task in follow-up activities is also important.

2) It is important that students be given opportunities to practice skills so that the teacher may monitor their understanding, provide feedback, and adjust her teaching accordingly. Accomplishing this goal in first-grade reading groups involved several distinct skills. Preventing misbehavior that could distract the teacher and other students was important, as was efficient use of time within the group. (For example, teachers who were ready to start the lesson immediately, without needing to organize materials or to correct students, had higher scores.) Students should be selected to respond in an efficient way, in order to insure that everyone gets a chance to practice. Within first-grade reading groups, ordered selection was often an effective technique to accomplish this purpose. At least in the setting we studied, it was important that most (although not all) of the questions asked to the students resulted in correct answers. For students at this age who are learning a skill that must become automatic, a low rate of errors may be necessary to supply enough correct practice of the skills in order to internalize them. For students at different ages and with different objectives, more or less difficult material may be appropriate. Indeed, some research has indicated that the optimal difficulty level may depend on several factors, including learner aptitude and motivation (Brophy and Evertson, 1976; Crawford, Note 11).

3) The teacher should provide as much information as is appropriate about the structure of the skills involved, rather than focusing only on memorized rules or labels. Positive results for several specific variables supported this principle: a) using overviews (which presumably help to set the stage for the steps to follow); b) using sustaining feedback after errors (which accomplishes two goals--the student receives practice at

256

successfully deriving the answer, and the sequential steps to the solution are made obvious when the teacher gives the clues to help break down the question); c) providing process feedback (so that an explanation of the steps is provided by the teacher); d) occasionally asking new questions (that generally are related to the preceding one, so that the same student is following up on his or her earlier answers); e) being specific when offering praise and criticism (high-lighting behaviors of interest).

Teachers who use such techniques some of the time may be helping the students focus on the relevant aspects of the problem solving process. However, the data indicate that these techniques are often not appropriate when they are overused and when they could interrupt the pace of the lesson (with these data, especially with oral reading turns). When the pace is broken, the immediate objective of the lesson may be lost (as, for example, when the purpose is to comprehend the passage), or the other students' attention may wander, in which case time-consuming behavior problems may result. Such techniques as listed above will probably be more or less appropriate in any setting, depending on how easily the teacher can work them into the lesson and how important it is that students see the sequence behind solutions.

4) Underlying all of the other principles, and making it possible to implement them in instruction, must be good overall classroom management. In the well-managed classroom, students follow efficient routines for accomplishing daily tasks, and there is a calm, pleasant environment in which all students may work without distraction. This means that the teacher prevents behavior problems, and that the students concentrate on the academic tasks. In this study, evidence of good management in the

more effective teachers' classes was found for such variables as transition time and fewer behavior corrections. Indeed, these teachers had implemented an effective management structure that made it possible to accomplish other goals, such as extended reading groups, more content covered, and greater student participation. Much information is being generated currently about how teachers may establish smoothly running classrooms. The growing body of research now makes it possible to attempt to change teacher managerial behaviors through efforts such as the present study (Kounin, 1970; Brophy and Putnam, in press; Evertson and Anderson, Note 9; Anderson and Evertson, Note 10).

The findings from the First-grade Study are in line with the concept of "direct instruction" espoused by Rosenshine and others (1976; Rosenshine and Berliner, 1978). These data provide specific information about strategies for implementing direct instruction in one setting--reading groups in first grade. Research conducted in other settings will help to distinguish further the specific requirements of the setting from the general principles that apply to a large population of learners and objectives. Through this kind of systematic approach to identifying process-product relationships, and then through experimental efforts such as the First-grade Study and others discussed in the next section, the knowledge base about effective teaching will become a practical and valid source of information and direction for the training and continued education of teachers.

Chapter 6: Suggestions for Future Experimental Studies of Teaching

The First-grade Study was successful in meeting many of its objectives. Teachers in the treatment group did use parts of the instructional model, and at least part of the achievement differences can be related to the model. As described in Chapter 5, the process-outcome relationships supported many of the instructional principles around which the treatment was built.

On a broader level, we believe that the results of this study support the efficacy of process-product research in the classroom, and the utility of experimental studies that attempt to modify teachers' behavior while substantiating earlier research findings. Through such efforts, classroom researchers are able to see findings translated into practice while validating the original results and refining prescriptive principles to include the "exceptions to the rule."

In order for validation and refinement to occur, the process of applying research findings must be conducted and evaluated carefully. This means that an experimental paradigm must be developed for use by classroom researchers that provides both rigor and relevance. Although an experimental approach is certainly not new to educational researchers, most experimental work with instructional variables has relied on artificial settings or controls within classrooms that isolated variables of interest. Such an approach allows experimental rigor, but lacks the flavor of the real-world setting of the classroom. The results are not generalizable, and therefore are not usable by practitioners.

However, using the natural setting of the classroom means that an experimenter must deal with many real-world constraints outside his or her control. Because of the complexity of a classroom, an experimental study that does not control or account for important factors may raise more questions than it answers.

This, then, is the dilemma facing the classroom researcher who wishes to conduct experimental studies: there must be a compromise between rigor and relevance. The experimenter must maintain the "ecological validity" of the setting and also maintain control of or account for factors that affect teacher behaviors and student outcome.

The prospect of juggling these two priorities is not as grim as it might seem at first glance. At this point, three large-scale studies have been reasonably successful in attaining this compromise and influencing both teacher behaviors and student outcomes. (In addition to the First-grade Study, see Good and Grouws, Note 8, and the work of Gage, Crawford, and associates, Note 12.) By drawing on the experiences of these researchers, several suggestions can be made about experimental design. Brophy (Note 13) made some of these as reviewed below, and other suggestions became apparent while reviewing the data from the First-grade Study. Therefore, the following sections offer a retrospective view of how to design an experimental study of effective teaching.

Maintaining the Ecological Validity of the Classroom in an Experimental Study

The purpose of an experimental study in a classroom setting is to influence teachers to adopt the behaviors and principles defined in a treatment. Generally, the treatment is a compilation of earlier research on effective teaching and typically suggest strategies for dealing with routine situations and decisions. This focus on the typical requirements of classroom life is not fortuitous. Research on effective teaching has demonstrated that strategies for dealing with "typical" situations differentiate more and less effective teachers. Examples are: arranging students for instruction, sequencing and pacing lessons, selecting students to answer, responding to student answers, scheduling instructional activities, and creating and maintaining order. Research has demonstrated that teachers' performance of such tasks has effects on student outcomes. Since teachers expend most of their time and energy in daily, routine decisions and tasks, efforts at improving their performance in these areas is likely to have payoff by affecting those student outcomes most closely associated with the teaching tasks.

The Content of the Treatment

Therefore, one of the most important considerations in the design of an experimental study is the content of the treatment--what it is that the teacher is supposed to do. If teachers are to implement the suggestions in the course of their regular instruction, the treatment must convince them to do so as effectively as possible. Doyle and Ponder (1977) suggested that the "practicality ethic" determines whether or not teachers will effectively use advice. Practicality depends on three qualities: it must be operational (easily translated into behavior); it must be congruent with the teacher's own role definition; and it must be efficient in terms of the teacher's cost and time. These three suggestions imply that a treatment will be implemented most easily when it is specific in terms of routine teacher behaviors, and when it provides a rationale that effectively relates the behaviors to the teacher's goals of instructing the students. Also, the behaviors must not make extensive demands on the teacher's time and energy, or at least none that are not compensated for in some way. (Of course, an adequate rationale that convinces teachers of the validity of the treatment often will be sufficient to justify extra effort on their part.)

Is is possible to influence more drastic change in teachers' behaviors, but it seems obvious that treatments that do not offer extensive support and compensation in terms of time and energy will not lead to adoption of novel or complex behaviors. Therefore, suggestions that are difficult to incorporate into the teacher's daily routine are not likely to be implemented easily. However, as the present study and the other experimental classroom studies have demonstrated, changes in student outcome may be influenced by treatments that are limited to familiar teaching tasks and that focus on the principles defining successful strategies for fulfilling those tasks. Such treatments offer more hope for effecting improvements in instruction and learning than more radical innovations that do not have an adequate research basis (such as changing to individualized assignments for each child or movement to open classrooms, for example).

In summary of this first point, the content of a successful treatment must reflect the teacher's understanding of the daily demands of the classroom in order for him or her to implement it thoroughly. Evaluation of a treatment depends on its being implemented, and so the classroom researcher must develop treatments that are likely to be implemented if the hypotheses are to be tested. The research base on teaching effectiveness contains many data that meet this requirement. Studies based on these data have been successful in influencing student outcome through teacher utilization of the suggestions about routine teaching tasks.

Clustering of Specific Suggestions

However, if we require that a treatment be operational and reasonable in terms of routine classroom tasks, another problem immediately arises. The treatment must be specific enough to allow translation into actual behavior, but there are inherent difficulties with very specific advice. Specific suggestions must be imbedded within larger principles, since no isolated behavior can be appropriate all of the time. However, it is not possible or desirable to list all possible situations and the variables defining appropriateness in them.

Therefore, a more reasonable approach to treatment design is to identify general principles of effective instruction and to cluster specific strategies under each. The specific strategies then serve as examples of the larger principle, and may themselves be organized around familiar teaching tasks and settings. When treatments are communicated in this way, the suggestions about particular techniques are imbedded in contexts and supported by rationales that can be discussed in common sense terms.

The following lists suggest categories that might be useful in organizing a treatment. Included are general principles of effective instruction that are supported by research; settings and contexts in which teachers function, and which may require different techniques according to the demands of the setting; and the teaching tasks that are regularly performed in almost any setting. Specific suggestions about techniques (such as use of ordered turns and sustaining feedback) can be organized by teaching tasks, with their appropriate use being a function of the setting demands and constraints. The rationale underlying their use is provided by the larger principle. The following lists are not intended to be complete, but only to serve as examples. There are, of course, other frameworks for discussing classroom activities and teaching strategies. For another example, see Good and Power (1976).

Principles of Effective Instruction

1. Achievement of content will depend on the learner's engaged time on tasks related to that content.

1a. Students need sufficient opportunities to practice components of complex skills and receive feedback on their performance.

2. In order for a teacher to be an effective instructor, there must be good classroom management that prevents many problems and minimizes distractions and interruptions of task-related activities.

Settings and Contexts in which Teaching Tasks are Performed

Settings which teachers may choose:

1. Arrangement of students (whole-class, small-group, individual students working with the teacher, individual students working independently).
2. Nature of group (members very similar or dissimilar in abilities and backgrounds).
3. Lessons with different objectives (e.g., presentation of new material vs. practice of new material vs. review of old material: learning of facts and labels vs. learning skills that require coordination and integration of facts and principles).

Settings which are not usually chosen by the teacher:

4. Student characteristics (classes with students who are unusually

low or high in ability and readiness skills, motivation).

5. Class size (large vs. small).
6. Time of year (beginning of year vs. later in year).
7. Time of day (morning vs. afternoon).
8. Type of classroom (large vs. small, open vs. self-contained).

Teaching Tasks

1. Arranging students for instruction, scheduling instructional activities, starting and ending transitions between activities.
2. Explaining new content, assigning work, selection of student activities for practice of new content.
3. Selecting students to answer questions, asking questions, responding to student answers.
4. Establishing and presenting rules and procedures, correcting misbehaviors, monitoring student behavior.

It is probably not possible (or desirable) to create treatments that include suggestions under all or even several combinations of categories, unless the treatment program is to be a long-term effort (such as a sequence of skills presented during the years of internship and induction into teaching, or an ongoing inservice program). Even if a treatment addresses only a few combinations of principles, settings, and tasks, an overall framework is valuable because it lends necessary perspective to an examination of the separate components of effective teaching.

As an example of how such categories could be used to cluster points in a treatment, consider the suggestions in the First-grade Study treatment about selection and feedback techniques.

PRINCIPLE: Students need sufficient opportunities to practice

265

components of complex skills and receive feedback on their performance.

SETTING: The teacher is responsible for presenting new phonetics rules to first-grade students. The students have been placed in a small group. The teacher's purposes are to present the new rules and to determine that the students can apply it with the words she will use in the lesson.

TASKS: In order to determine that the students have learned the rule, the teacher will ask them questions and correct them if they make errors.

SUGGESTED TECHNIQUE: Selection. Select students by going around the group in order. This will insure that there is equal opportunity for practice, and it will help control problems with call outs and over-enthusiastic students dominating the action.
CAUTION: Be careful that the students are not "tuning out" when it is not their turn. If this occurs, then the overall objective of maximizing individual attention to and practice with the skills is not being met.

SUGGESTED TECHNIQUE: Feedback to incorrect answers. Help the child derive the correct answer himself by asking a series of questions that simplify the problem and focus the child on the sequence to follow. If this is not successful, give the child the answer and move on. Occasionally, describe the sequence that must be followed to derive the answer.

CAUTION: Be careful that the attention of other students is not lost because of too much time spent in working with one student. Don't let the overall pace suffer. Also be careful that the simplified questions do not become "pointless pumping", which embarrass the student, thus defeating the purpose of providing successful practice with the skill.

Obviously, when the purpose of the lesson or the setting is very different, the tasks will remain the same, although the techniques suggested above will be appropriate less often. For example, if the teacher chose to present a fast-paced drill to check for rapid recognition of sight vocabulary words, then frequent use of sustaining feedback would disrupt the pace. When students are reading aloud for the purpose of story comprehension, sustaining feedback would again be inappropriate much of the time, because the continuity of the passage would be lost while the teacher and student worked on a single word. If the use of ordered turns resulted in less attention than desired, the teacher might want to vary the selection techniques. The immediate objectives are to distribute practice opportunities fairly, and minimize problems with over-enthusiastic students who might "upstage" more reticent students who also need to interact with the teacher. Many teachers have found that ordered turns are useful in achieving these objectives in this setting, but other types of selection may fulfill the same purposes.

The purpose of presenting suggestions in this way is to help teachers choose among alternative strategies according to a rationale that defines appropriateness. That is, the purpose of "treating" teachers is not to force them to use a specific technique every time it is possible, but instead to optimize use of several techniques or strategies--using them when appropriate and avoiding their use when inappropriate. Training teachers to make such decisions requires a conceptual framework with which they may examine their classroom tasks along with a rationale that explains why certain strategies are or are not effective in various settings.

The implications of this for treatment design are that specific suggestions must be placed within a meaningful framework. This is an important part of meeting the first objective of classroom experimental design: maintaining ecological validity. Teachers daily make thousands of decisions about the tasks listed above and no treatment, no matter how thorough, will replace the teacher's use of his or her own best judgment.

Maintaining Experimental Control in a Classroom Setting

Although the most difficult part of designing an experimental program will be the development of the treatment, it is equally important to set up the study to maintain experimental control. Otherwise, the content of the treatment can not be evaluated. In the traditional sense, "experimental control" implies elimination of any contaminating or modifying influences, so that the treatment is the only possible explanation for differences between experimental and control groups. It is not possible to maintain control in this sense in classroom studies. Therefore, it is important to establish as much control as possible through initial selection of subjects and matching of groups, and then to closely monitor other factors that may have important effects on outcomes. In these ways, control is exerted either through initial assignment or with statistical techniques. Examples of influential factors that should be controlled are given in the points below, which discuss some procedures that should be considered by classroom researchers who conduct experimental studies.

1. Information on classroom processes and characteristics should be collected before the treatment is applied. In this way, one may insure that treatment and control groups are equivalent in important ways before the treatment is introduced. Measurement of preexisting differences may allow formation of equivalent treatment groups, especially if a larger initial sample is available than will be used in the actual data collection. Some of the dimensions on which classes will differ and which have been shown to influence student outcome are student ability level and background, teacher managerial effectiveness, school policies on allocation of time, curriculum and materials, and pacing of content.

These pretreatment measures of the treatment and control groups were not taken in the First-grade Study. Consequently, some of the group differences on classroom process scores could not be explained. This meant that many questions were left unanswered about possible expectancy effects and preexisting differences, and therefore, the full impact of the treatment could not be determined.

2. Data should be collected to describe the actual level of implementation of behaviors and principles described in the treatment in both

the treatment and control groups. Other research has demonstrated that the effects of an innovation may be most easily explained by measuring the actual implementation, rather than by assuming differences between the treatment and control groups (Hall and Loucks, 1977). Not all treatment teachers will implement to the same degree, and some control teachers will demonstrate high natural levels of use of the techniques or principles being examined. This latter possibility is especially likely when the treatment is based on studies of effective teachers. Some teachers in a control group will already be using strategies that are associated with effective teaching. If the treatment is to be fully evaluated, then the actual extent of use by all teachers must be known.

Researchers should also examine alternative strategies to those described in the treatment, and should measure other factors that might influence outcome, as described under the first point above. Even if these were measured at the beginning of the study in order to form equivalent groups, it is important to note changes and development over time. (For example, management strategies that are effective at the beginning of the year may not be equally effective later in the year.) Also, some important variables can only be evaluated at the end of the study or school year, such as total amount of content covered.

3. Measures of short-term outcomes should also be taken. In this way, short-term effects of the treatment may be detected. This will be important if the treatment has an effect that is different from that expected by the researcher. For example, in the First-grade Study, we suggested that transitions could be made more effective by using standard signals to notify the students. We found that most teachers, including

control teachers, did this, and so there was no group difference in implementation of that principle. Neither was there a relationship with outcome due to the restricted variance. (This did not indicate that the principle was invalid, but that it represented a very common behavior.) However, there were differences between the groups for the efficiency of transitions (with the treatment group having shorter, smoother ones.) Since there were no measures of preexisting differences, we can not rule out the possibility that the treatment teachers had smoother transitions to begin with. However, an alternative explanation is that the treatment may have caused the teachers to focus on their transitions and work harder to make them more efficient, even though the strategies given in the treatment were obviously not ones that made a difference. However, the short-term outcome measure (efficiency of transitions), provided support for an alternative explanation of treatment effects: increasing awareness of need leads to improved behavior, regardless of specific suggestions.

Another advantage of measuring short-term outcomes is that their relationship to long-term outcomes (such as student achievement) will make it clearer why the long-term outcomes are related to the specific teacher behaviors (that presumably led to the short-term outcomes). Also, by including immediate effects of teacher behaviors, the researcher may be better able to distinguish true effects of treatment content from Hawthorne effects (discussed in the next point) that might influence long-term outcomes.

4. There is always the possibility that treatment teachers will outperform those in a control group simply because of the special attention and/or the expectations of better performance communicated to them

(e.g., "We think that this treatment will make you a better teacher"). Therefore, any apparent treatment effect must be related to the actual content of the treatment, rather than to heightened expectations for success with the accompanying extra effort to fulfill the prophecy. Otherwise, the principles in the treatment have not really been tested, and the possibility of Hawthorne or expectancy effects must be considered.

However, each of the experimental studies have utilized designs that allowed the researchers to sort out the effects of treatment content from other factors, although none of the studies excluded all possible factors. Good and Grouws (Note 8) tried to heighten an expectancy effect by informing his control group that their performance would be evaluated after a few months, at which time they would be given feedback along with the treatment that was supplied to the other teachers. Although the student achievement in this group was improved, there was a greater change in classes where teachers had been given the specific treatment. It was concluded that the treatment had a positive effect that could be distinguished from the effects of the encouragement given the control group. Gage, Crawford, and associates (Note 12) utilized two treatment groups with differing levels of information provided to them. The minimal group received the written information and one presentation by the experimenters (similar to the First-grade Study's treatment), and the maximal treatment group received feedback and extra encouragement. Differences between these two treatment groups can be interpreted as effects of treatment strength. In the First-grade Study, two treatment groups were also utilized to ascertain the effects of observation on outcome when a treatment was given. It was expected that the observed group would follow the treatment model more closely and might try harder to do well because of the frequent observation. (This design would have been more effective if a comparable unobserved-control group had also been included for examination of student outcomes.)

Although it has not been utilized, there is another design that could serve to separate Hawthorne from content of treatment effects. This would include two treatments, focusing on different but important aspects of teaching. Each group would receive the same amount of information regarding its own treatment and the same expectations for success. Each would be observed to determine implementation of its own treatment, as well as behavior in other aspects of teaching. If the treatment content was truly responsible for changes in behavior and outcomes, then the two groups would differ for those behaviors emphasized in their particular treatments. For example, one treatment could focus on reading group management, while another focused on managing whole class lessons. In the first treatment, one would expect to find improved teaching in reading groups, but not in whole class instruction unless the treatment was having a more generalized effect. In the second group, the most evident changes in behavior should be in the area of whole class lessons. However, if both groups showed improvement in several areas of teaching, and the teachers were not necessarily stronger in the area specified by their treatment, then one could conclude that Hawthorne or expectancy effects may have been operating. An additional advantage of this design would be the opportunity to examine generalized use of specific suggestions in other settings.

In summary, by utilizing designs that vary the expectations conveyed and the amount and type of information given to the teachers, it may be possible to separate the effects of actual treatment content from other effects on teacher performance. An additional advantage of such designs is that future decisions can be made on the basis of cost-effectiveness. For example, if a minimal treatment is as effective as a more extensive (and expensive) treatment in leading to a desired outcome, then the minimal treatment could be more easily exported to other users, such as staff developers, teacher educators, etc.

5. It is also desirable to determine any effects of teacher characteristics and ability. It is possible, and indeed likely, that one level of treatment may be effective with one teacher and not with another because of the teachers' experience, aptitude, and entering skills. Although no attempts were made to measure such teacher traits in the First-grade Study, the other two studies included such measures, and both found that teacher characteristics were important predictors of student outcome, either in interaction with or independent of the treatment.

6. It is also important to consider setting characteristics. As was discussed in the first section of this chapter, the content of the treatment will be valid only as long as it recognizes the many contexts in which teachers must act. These setting variables must be considered in evaluating the implementation and effects of a treatment. For example, in the First-grade Study, there were important differences in implementation of certain strategies depending in the kind of lesson being taught--whether it was a question-answer sesson, a lesson involving writing in workbooks or worksheets, or a lesson involving oral reading. Although the treatment did not mention the different requirements of these types of lessons, the teachers utilized the principles differently in them. In many cases,

when implementation scores were examined, there were significant context-treatment interactions: the treatment group was different from the control group in one lesson context but not in another (Anderson, et al, Note 5). These interactions were sensible: generally the treatment teachers did not utilize a technique that would have been inappropriate for the pacing requirements of the setting, although they would use that technique more than the control teachers in a setting in which it was appropriate and useful. It is important to note that, although some contextual distinctions were built into the treatment (especially about types of feedback), lesson format was not. However, the observation instrument included much information not clearly linked to the treatment in the hopes that other key factors would become obvious. In this case, lesson context emerged as an important consideration, especially the difference between reading turns and questions asked out of turn.

Conclusion

Obviously, it is difficult to plan and conduct a study that includes all of the above suggestions. However, the points discussed should be considered in designing and evaluating a treatment, both in terms of its effects on teacher behaviors as well as on short- and long-term student outcomes.

The last decade of classroom research has provided a body of information that makes it possible to create relevant treatment models that can be used by teachers. More sophisticated ways of organizing and analyzing several types of data, including multiple regression techniques, make it possible to examine many simultaneous effects on classrooms. By taking advantage of knowledge in each of these areas, classroom researchers who are determined to see their results translated into practice can do so through treatments and experimental designs characterized by the points described above. Such designs, along with continued descriptive and correlational research, will yield a more complete picture of what works in classrooms, when it works, why it works, and how to make it happen.

Reference Notes

1. Brophy, J., & Evertson, C. *The Texas teacher effectiveness project: Presentation of non-linear relationships and summary discussion*. Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1974. (ERIC ED 099 345)
2. Brophy, J., Anderson, L., Greenhalgh, C., Ogden, J., & Selig, H. *An instructional model for first grade reading groups* (Report No. 4043). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1976.
3. Medley, D. M. *Teacher competence and teacher effectiveness: A review of process-product research*. Washington, D. C.: American Association of Colleges for Teacher Education, 1977.
4. Brophy, J., Mahaffey, L., Greenhalgh, C., Ogden, J., & Selig, H. *Coding system for the first grade reading group study* (Report No. 4013). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1976.
5. Anderson, L., Morgan, R., Evertson, C., & Brophy, J. *Context effects and stability of teacher behaviors in an experimental study of first grade reading groups* (Report No. 4091). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1978.
6. Veldman, D., & Linsley, T. *Programs JHI and CURVE*. *Special purpose program using PRIME library routines to solve and compare linear models*. Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1977.

7. Martin, J., Anderson, L., & Veldman, D. *A methodological approach to the unit of analysis problem in classroom research* (Paper in progress). Austin, Texas: University of Texas, Research and Development Center for Teacher Education.

8. Good, T. L., & Grouws, D. *Experimental study of mathematics instruction in elementary schools*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

9. Evertson, C. M., & Anderson, L. M. *Interim progress report*: *The classroom organization study* (Report No. 6002). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1978.

10. Anderson, L. M., & Evertson, C. M. *Classroom organization at the beginning of school*: *Two case studies* (Report No. 6003). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1978.

11. Crawford, J. *Relationships between grade point average, response mode, difficulty level, and learning* (Report No. 4021). Austin, Texas: University of Texas, Research and Development Center for Teacher Education, 1975.

12. *An experiment on teacher effectiveness and parent-assisted instruction in the third grade*. Stanford, California: Stanford University, Center for Educational Research, Program on Teaching Effectiveness, 1978.

13. Brophy, J. *Training teachers in experiments*: *Considerations related to nonlinearity and context effects*. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977.

References

Ausubel, D. P. *The psychology of meaningful verbal learning*. New York: Grune and Stratton, 1963.

Becker, W., & Engelmann, S. The direct instruction model. In R. Rhine (Ed.), *Encouraging change in America's schools: A decade of experimentation*. New York: Academic Press, in press.

Blank, M. *Teaching learning in the preschool: A dialogue approach*. Columbus, Ohio: Charles E. Merrill, 1973.

Blank, M., & Solomon, F. A tutorial language program to develop abstract thinking in socially disadvantaged preschool children. *Child Development*, 1968, *39*, 377-389.

Bloom, B. *Human characteristics and school learning*. New York, New York: McGraw Hill, 1976.

Borich, G. D. (Ed.). *The appraisal of teaching: Concepts and processes*. Reading, Massachusetts: Addison-Wesley, 1977.

Brophy, J., & Evertson, C. *Learning from teaching: A developmental perspective*. Boston, Massachusetts: Allyn and Bacon, Inc., 1976.

Brophy, J., & Good, T. *Teacher-student relationships: Courses and consequences*. New York: Holt, Rinehart and Winston, 1974.

Brophy, J., & Putnam, O. Classroom management in the elementary grades: A literature review. In D. Duke (Ed.), *Classroom management: Seventy-eighth yearbook of the National Society for the Study of Education*. Chicago, Illinois: University of Chicago Press, in press.

Chall, J. *Learning to read: The great debate*. New York: McGraw-Hill, 1967.

Doyle, W., & Ponder, G. The practicality ethic in teacher decision-making. Interchange, 1977, 8, 1-12.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart and Winston, 1974.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan Achievement Tests, Primary 1 (teacher's directions): Forms F, G, and H. New York, New York: Harcourt, Brace, Jovanovich, Inc., 1970.

Good, T., & Power, C. Designing successful classroom environments for different types of students. Curriculum Studies, 1976, 8, 45-60.

Good, T. L., Grouws, D. A., & Beckerman, T. M. Curriculum pacing: Some empirical data in methematics. Curriculum Studies, in press.

Groisser, P. How to use the fine art of questioning. New York, New York: Teachers' Practical Press, 1964.

Hall, G. E., & Loucks, S. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14, 263-276.

Hildreth, G. H., McGauvran, M. E., & Griffiths, N. L. Metropolitan Readiness Tests: Forms A and B. New York, New York: Harcourt, Brace and World, Inc., 1969.

Kleinfeld, J. Effective teachers of Eskimo and Indian students. School Review, 1975, 83, 301-344.

Kounin, J. S. Discipline and group management in classrooms. New York: Holt, Rinehart and Winston, 1970.

Loughlin, R. On questioning. Educational Forum, 1961, 25, 481-482.

Rosenshine, B. V. Classroom instruction. In N. L. Gage (Ed.), The psychology of teaching methods: Seventy-fifth yearbook of the National Society for the Study of Education (Part I). Chicago, Illinois: University of Chicago Press, 1976.

Rosenshine, B. V., & Berliner, D. C. Academic engaged time. British Journal of Teacher Education, 1978, 4, 3-16.

Rosenshine, B. V., & Furst, N. The use of direct observation to study teaching. In R. Travers (Ed.), Second handbook of research on teaching. Chicago, Illinois: Rand McNally, 1973.

Southwest Educational Development Laboratory. Bilingual kindergarten program, inservice manual (Vol. I). Austin, Texas: National Educational Laboratory Publishers, Inc., 1973.

Ward, J. H., & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice Hall, 1973.

Weinstein, R. Reading group membership in first grade: Teacher behaviors and pupil experience over time. Journal of Educational Psychology, 1976, 68, 103-116.

Appendix A

AN INSTRUCTIONAL MODEL FOR FIRST GRADE READING GROUPS

To the teachers:

This is a description of a system for small group instruction of young children (in this case for first graders in reading groups.) It does not discuss content or materials, but it provides guidelines for teacher management of reading group instruction. It is hoped that the systematic use of these principles will improve the planning and conduct of reading group sessions and benefit the children. A major underlying rationale for the system is that each child should receive as much individualized instruction as is possible in a group setting.

The principles discussed in the following pages flow from both experience and research involving teachers and young children in small groups. By combining them into an organized system to be used in the classroom, much more information can be gained about how to best teach small groups of young children.

The purposes of this project are to bring these principles together into a workable system and to teach teachers to use them if they are not already doing so. (You may recognize many of the principles as techniques which you already use.) After asking you to incorporate these suggestions into your teaching, we will examine the results in order to further evaluate the system and the principles. These findings then could be used in teacher education and teacher inservice programs if they show that certain techniques make a difference in children's learning.

You probably will find that many of the principles are more applicable at one time than another, depending upon which children you are teaching and what kind of lesson you are presenting. We have tried to provide a general overview that can be adapted to the many different lessons and types of children with which the first grade reading teacher must deal. There is a special emphasis on dealing with shy, impulsive, and inattentive children and problems such as wrong answers and failures to respond. It is hoped that dealing with such situations in the suggested ways will make reading a more pleasant and productive experience for both the teacher and the children.

Your role in the study is central, because application of these principles involves teacher judgment based upon knowledge of individual children's needs as well as a feel for the group's needs. Specific examples have been provided but are not meant to serve as absolute prescriptions. Rather, we ask that the teacher learn the general principles, and then use them according to her best judgment about the situation and the children involved.

Please read the material and study it until you are comfortable with it and feel that you could conduct a reading group accordingly. A meeting will be scheduled at your school to discuss any questions and comments. There will be a short test administered at the end of the meeting to assure understanding of the principles. Any areas of misunderstanding that show up on the test will be discussed again, so that both you as a teacher and we as researchers can reach mutual satisfaction and agreement about procedures.

INTRODUCTION

The instructional model is based on two general principles concerning children's learning in small groups:

1. It is desirable to have a balance between a.) an efficient group structure in which the pace is rapid enough to maintain interest and attention, and b.) a group structure which helps the teacher to make sure that learning is taking place for every child.

2. It also is desirable that children learn to respond to every teacher question, but without feeling anxious about having to make a response.

Accomplishing either of these goals requires teacher judgment at many points in the lesson. How fast should questions be paced to keep attention and yet not lose anyone? How long can you wait for a response from an individual without losing the attention of the rest of the group? When should you end a child's response opportunity if he might know the answer but seems afraid to say anything? How long should a child be urged to respond before such encouragement creates embarrassment and anxiety?

Specific answers to these questions cannot be prepared in advance, since the situation is different for every child and every question. However, the system of principles outlined below can be used as a framework within which the teacher, who knows the children, can make decisions.

The principles are presented below in a brief list. In the next section they are discussed in greater detail, along with the rationales and background information related to them. The system is divided into two major components: 1) organization and mangement; and 2) teacher responses to children's answers.

In order to avoid confusion of pronouns, "she" will be used to refer to the teacher and "he" will refer to the student.

OVERVIEW OF THE PRINCIPLES

I. ORGANIZATION AND MANAGEMENT

GETTING THE CHILDREN'S ATTENTION

1. The teacher gets everyone's attention before starting the lesson.

2. The children sit with their backs to the rest of the class while the teacher faces the class.

INTRODUCING THE LESSON

3. The teacher introduces the lesson with a brief overview.

4. The teacher presents new words clearly.

5. After presenting new words, the teacher has the children repeat them.

6. A demonstration or explanation precedes the children's attempts to do the work.

CALLING ON CHILDREN

7. The teacher should work with one child at a time, so that everyone is checked and receives feedback.

8. The teacher should call on children in order rather than randomly.

9. Occasionally the teacher should question a child about another child's response (to keep everyone alert).

10. The teacher should minimize calling on volunteers.

11. The teacher should discourage call outs and should emphasize that each child is responsible for the question asked of him.

12. The teacher should avoid rhetorical questions, answering her own questions, or repeating questions. These confuse the children.

MEETING INDIVIDUAL LEARNING NEEDS WITHIN THE GROUP

13. At some point, the teacher must decide if the whole group can meet the lesson's objectives. If she decides they can, she should hold the group together, making sure that everyone masters each step before moving on to the next step.

14. If the teacher decides that everyone cannot meet the objective, the students who can do so should be taught through to the end and then dismissed, so that the teacher can spend more time with the other children.

15. An exception to the above occurs when the teacher wants to use a student who has mastered the objective as a model for the others. Here, she may retain one or more such students in the group in order to carry on a dialogue.

16. If some of the children do not succeed in meeting the objectives before lesson time is up, arrangements should be made for extra tutorial help.

II. RESPONDING TO CHILDREN'S ANSWERS

The teacher's feedback to children's answers depends on 1) the type of question (whether it requires memory or reasoning), 2) the pace of questioning (whether rapid for drill or slower for more thoughtful questions), and 3) the child's answer (correct, incorrect, "I don't know," or no response).

WHEN THE CHILD DOES NOT RESPOND

17. After asking a question, the teacher waits for the child to respond and also sees that other children wait and do not call out answers. During rapid pacing, she waits a few seconds and gives the answer. During the more slowly paced parts of the lesson, the teacher should wait for an answer as long as she feels that the child is thinking and will answer,

but not so long as to embarass the child or lose the other children's attention.

If the child does not respond within a reasonable time, the teacher should indicate that some response is expected by probing ("Do you know?"). She should then simplify (see #19) according to the type of question.

WHEN THE CHILD'S ANSWER IS INCORRECT

18. The teacher should indicate that the answer is wrong, and then follow simplification procedures outlined below for the two types of questions.

SIMPLIFICATION PROCEDURES

19. The appropriate simplification procedure is determined by the type of question.

a. If the question deals with factual knowledge that cannot be reasoned out, the teacher should give the answer to the child and then move on.

b. If the question is one that the child could reason out with help, the teacher should provide clues or simplify the question. If the clues still do not help the child, he should be given the answer. The teacher should never ask another child to supply the answer.

WHEN THE CHILD IS CORRECT

20. The teacher should acknowledge the correctness, and make sure that everyone else heard and understood the answer.

PRAISE AND CRITICISM

21. Praise is important but should not be used indiscriminately. Praise thinking and effort more than just getting the answer, and make praise as

specific and individual as possible.

22. Criticism should also be as specific as possible and should include specification of desirable or correct alternatives.

PART I: ORGANIZATION AND MANGEMENT

GETTING THE CHILDREN'S ATTENTION (Principles 1 and 2)

It is important to catch and maintain the children's attention at the beginning of the lesson.

1. The teacher should use a standard and predictable signal to get the children's attention. The use of this technique should lead to quicker and easier transitions with little time wasted in getting a group started. It is useful in two situations. The first is the transition from general class activities to the reading group (and alternate activities), and the second situation is getting everyone's attention when you actually begin the reading group.

A standard and predictable signal is one which the children can learn to recognize quickly because it is repeated daily with the same meaning. For example, the teacher might ring a bell every day to signify that it is time to move to the first reading group, or she might give a consistent verbal signal, such as "It's time for Tigers!" If the signal is clear and consistent, the children do not have to stop each time and decide what to do; they can respond quickly and automatically. The teacher should decide upon the signal early in the year and the children should be allowed to practice responding to the cue.

Once the children are in the group, the teacher should again use a consistent recognizable signal indicating that the lesson is about to begin and that the children should pay attention. For example, she could use a phrase every day such as "Attention, children." Again, by consistently using the same signal the children learn more easily when lessons are starting, and the teacher will spend less time getting the group organized.

In either situation, if some children do not respond to the signal, the teacher should remind them individually of what is expected. However, the teacher should be careful not to interrupt the appropriate behavior of the other children or reward the unattending child with too much attention (by shouting, scolding loudly, etc.). Instead, the teacher can tap or softly call the names of those who are not attending. If this still does not lead to the desired behavior, the teacher should quietly and quickly explain the meaning of the signal and make sure that the child follows through. It is important that the teacher remain consistent and firm in her demands that the children respond to the signal.

2. Once in the group, the children should be seated with their backs to the rest of the class while the teacher is facing the class. This is a preventive measure in that the children in the group are less likely to be distracted by other activities if they face the teacher and have their backs toward the rest of the class. Also, the teacher can supervise both the small group and the remainder of the class at the same time in this position.

290

INTRODUCING NEW MATERIAL (Principles 3,4,5, and 6)

The introduction should prepare the children for the lesson by getting their attention, teaching new material and new terms before applying them, and making sure that the children know what to do when they are asked to make responses.

3. *The introduction should contain an overview* of what is to come in order to mentally prepare the children for the presentation. This does not have to be elaborate (in fact, it should not be). A sentence or two will do, such as "Today we are going to learn about a sound the letter *a* makes." Such an overview should contain a statement about the content to be studied. It may or may not be expressed in terms of a behavioral objective (as in, "At the end of the lesson, you will know about two different sounds for *th*."). The overview may also mention something about the purpose of the lesson, especially if this is likely to motivate the students. For example, the teacher may say, "Today we are going to learn about words with two vowels side by side. When you know about this, you will be able to read a lot of new words that you could not before, like *boat*, and *seat*, and *sail*."

If the lesson will involve changing activities at some point (and especially if the teacher anticipates breaking up the group to work with certain children, as discussed in Principles 13 and 14), it may be useful to give the children a preview of the sequence of events. (for example, "Today we will talk about some words like *through*, *rough*, and *although*. They sound different but they look a lot alike. After we talk about them for awhile, some of you will go do workbook exercises, some of you will go to the listening center, and some of you will stay and talk with me."). This prepares the children for a future transition, and it also lets them know in advance that several different activities will take place. The children then will be expecting these directions when they are repeated later.

The purpose of overviews such as these is to prepare the children by helping them to organize their thinking and focus on the task at hand. It is important for them to learn that the world of school and its demands (specifically reading) is a reasonable and orderly one. One step toward helping them learn this is always to prepare them before making demands on them, and then to follow that plan accurately so that their expectations of the lesson are fulfilled.

4. It is also at the beginning of the lesson that new words and sounds should be presented to the children, so that they can use them later when they are reading or answering questions. Introduction of new words may be accomplished in several ways, depending upon the words, the children, and the teacher. Words that do not follow phonics rules and cannot be sounded out by the children should be said clearly by the teacher. Words that could be sounded out by the children may be presented as questions to them, along with whatever other clues may be helpful. Or, if the teacher prefers, these words also could be given to the children by her. The important thing is that these new words are spoken in some form at the beginning of the lesson. The teacher should see that they are distinctly pronounced and pointed to, explained, or otherwise focused upon. This technique lets the children know what to expect, so that they can read without anticipating totally unfamiliar words. They can be looking for them, and their learning of the words will be reinforced when they see them again.

5. When new words or sounds have been presented, the teacher should have the children repeat them until they can say them satisfactorily. Having them repeat the words or perhaps make up sentences with them gives them practice in reading and saying new words before they are called on to read them in context.

It is also a relatively easy task, so that children who are shy about responding will find this first demand less frightening than a more complicated question.

6. After moving into the lesson, but before asking the children to use new material or undertake new tasks, the teacher should present a demonstration and/or explanation of any new activity. A good explanation includes a step-by-step description of the processes involved , given in simple, clear language that the children can understand. The teacher should gear the explanation to the children's experience and level of understanding. For example, if the task was to find pictures whose names started with the same sound as the name of the letter just studied, the teacher might say, "Show me all the pictures whose names begin with the /b/ sound." But if the children had never performed this task before and were not familiar with the skills involved, this might be a poor explanation. A better approach would be to break the task into each step and explain sequentially.

For example, with a readiness group that had never before tried this task, the teacher might say (after an overview and presentation of the letter sound being studied), "First look at this letter, b. What sound does it make?" (Child responds.) "Say the sound to yourself. Now look at this picture. What is the name of the picture?"..."Say the name to yourself. What is the first sound in that name?....Is that the same sound as the sound of this letter?" If the children could answer each of these questions, the teacher could present the next picture with fewer questions. For example, she might say, "Now look at this picture and say the name to yourself. Listen to the first sound. Is the first sound the same as the sound of this letter?" Later, this could be further shortened to "Look at this picture and tell me if the first sound in its name is the same as the sound of this letter." Eventually, the children could be given the explanation presented first above ("Show me all the pictures whose names begin with the /b/ sound.").

Therefore, the criteria for a good demonstration or explanation involve checking the children for their level of understanding and then, if necessary, either expanding on the explanation by breaking it into steps or shortening it by leaving out a few steps at a time. The teacher's choice depends upon the children's responses at any point in the explanation.

After the explanation or demonstration, the teacher should move quickly to having the children do the task themselves. Children in this age range need concrete personal experience to learn concepts or skills, but they also need guidance to point out the most important features of the task. Having the children perform the task in the group not only helps them learn it, but also allows the teacher to check them for understanding of the instructions before they are released to work on their own or expected to respond correctly in a more rapidly paced group session.

CALLING ON CHILDREN (Principles 7,8,9,10,11, and 12)

Calling on children involves distributing response opportunities to individuals while at the same time keeping the entire group alert.

7. The teacher should work with one individual at a time in having the children practice the new skill or apply the new concept, making sure that everyone is checked and receives feedback during the lesson. (Feedback is discussed below in principles 16 through 21.) In this way, the teacher can monitor the progress of each group member. This means that excessive use of choral responses is not desirable.

8. The teacher should use a pattern (such as going from one end of the group to the other) for selecting children to take their turns reading in the group or answering questions (as opposed to calling on them randomly and unpredictably). For example, the teacher can start with the child to her immediate left, then the child to his left, and so on around the circle, questioning each child or asking for reading.

This is suggested because the children will always know when to expect a turn and will not feel anxiety about being called on unexpectedly. This is especially important with young children who feel uncertain about having to perform in school, and it also will help to control overeager students who frequently call out answers, wave their hands, and engage in attention-seeking activity because they think it will lead to a turn to read or respond. Both shy, non-responsive children and attention-seeking, overeager ones will know when their turn is coming and will not spend the rest of time feeling anxious or trying to get attention. The teacher must remain firm in her use of this procedure and not skip a shy child to yield to a handwaver, except in the situations discussed below.

9. In order to keep each member of the group alert and accountable at all times between turns, the teacher occasionally should question a child about a previous response from another child (for example, "Bill, how do you feel aobut John's idea?" or "Do you have anything to add to that?") Thus, each student should know that he may be called upon at any time, not just during his turn. He must therefore remain attentive and listen to the other children. However, two precautions should be taken when using this technique. First, when a child is questioned about another's response, the demand made on the child should be an easy one for him. That is, if he was listening and paying attention, he should be able to answer the question without difficulty. For some children, such demands are as simple as asking for a repetition or opinion. Other children might be asked to comment on the correctness of the answer or to expand upon it (but only if the teacher feels this is within their capabilities).

If harder demands are made than a child can fulfill with this type of questioning, the advantage of reducing anxiety by using a predetermined order (as discussed in principle 8) will be lost. If all of the out-of-turn questions are simple for each child, they will not learn to fear them. Instead, they will be rewarded for paying attention and listening, and they will get an extra opportunity to give a correct answer.

As a second precaution, the child should be helped to realize that the purpose of such questioning is to get his opinion or input, not to put down or "correct" another child. The teacher can serve as a model through her responses by treating wrong answers as a reason to teach, not to criticize.

The use of these two principles should create a desirable balance between predictability, which helps reduce anxiety and/or attention seeking behavior, and continuous alertness within the whole group.

10. Calling on volunteers should be primarily restricted to parts of the lesson in which children are contributing personal experiences or opinions. However, when the objective of the lesson is to teach some content or skill, it is important that every child be called upon and expected to respond. This can best be accomplished by using ordered turns and occasionally questioning children out of turn to keep them alert.

11. When call outs occur, the teacher should remind the child that everyone gets a turn and that he must wait until his turn to answer. It is important not to be overly critical, however, especially if the call out demonstrates enthusiasm in a child who usually does not exhibit it. Nevertheless, all children should learn that when one child is asked a question, he is responsible for the answer, and others are not to call out the answer or "help."

If a child persists in calling out despite repeated reminders, the teacher must determine why he is doing so. Her later reactions then are determined by the reason for the behavior. For example, if the calling out primarily seems to be to get attention, the teacher should make sure that her responses are not reinforcing the behavior by paying attention to it. Reminders can be delivered impersonally to the child, without looking at him or seeming to speak directly to him (thus not rewarding him with attention), or the teacher can totally ignore the student's call outs and only respond to answers given during his turn.

Another reason for repetitive call outs might be that the child is impulsive and has little self-control. In this case, the teacher may help the child become aware of his behavior so that he can begin to control it.

It is important that the teacher never accept a called-out answer. Call outs should be ignored or should result in a reminder that everyone is expected to wait his turn or raise his hand and be called on by the teacher. In contrast to this response to call outs, the teacher should be sure that she does respond positively to answers given during a child's turn. (In the case of a typically non-responsive child who does make a rare call-out, the teacher should not lose the opportunity to reinforce the response, while gently reminding the child if she can that it would be better to answer during his turn. Otherwise, however, the teacher should not accept the answers called out.)

The purpose of this principle is to not only help maintain control but to teach children to listen to others and not to interrupt.

12. The teacher should avoid rhetorical questions, asked for effect with no answer expected, or leading questions ("Wasn't that funny?"). Other questioning patterns to be avoided are answering her own questions ("Why did the farmer go to town? To buy a pig, of course!", without waiting for an answer) and repeating questions ("Why did the farmer go to town? What did he want to do? Why did he go?", again without waiting for an answer). These kinds of questions tend to confuse the children and will also make it more difficult to teach them that each teacher question demands an answer. When the children are always asked questions that can be dealt with and have sensible answers, they are more likely to form the attitude that school demands are reasonable and can always be answered eventually.

When rhetorical or leading questions are asked frequently, the children may learn that an answer is not expected or that it can be figured out from the tone of the teacher's voice (as in "Wasn't that funny?" or "Don't you feel sorry for poor old Nobbin?"). If the teacher frequently answers her own questions without pausing for an answer, the children may be confused and not see the connection between the different expressions of the same question. To them, the above example ("Why did the farmer. . .? What did he want? Why did he go?") might appear to be several questions at once, which could confuse some first-graders.

Instead, good questions for this age group should be short enough for the children to hold in their memory while thinking about the answer. They should elicit some mental activity beyond second-guessing the teacher (responding to her tone of voice rather than the content of the question), and they should have answers which make sense to a young child who cannot think abstractly or juggle too many concepts at the same time. By consistent use of reasonable questions, the teacher can help promote in her students the idea that school tasks are reasonable and within their capabilities.

MEETING INDIVIDUAL LEARNING NEEDS WITHIN THE GROUP (Principles 13,14,15, and 16)

Meeting each individual's learning needs may involve breaking up the group, using another child as a model, and arranging for tutorial help.

13. At some point during the lesson, the teacher must make a fundamental decision about whether the group as a whole can or cannot meet the lesson's objectives. If there are large individual differences in the rate of learning, keeping the group together might mean spending too much time with those who are having difficulty. When this is not the case and the group as a whole can meet the objectives, the teacher should keep the group together, concentrating her attention at each step on the slowest members, working with them until they master the step before proceeding to the next one. In this fashion, all of the children will achieve at least the minimal objectives of the lesson.

14. If the teacher decides that the group as a whole cannot reach the objectives at the same time because of large individual differences in comprehension of the material, she should proceed differently. Those students who already know the objectives or who are learning rapidly and easily should be taught through to the end of the lesson and then dismissed from the group to work independently or engage in some approved self-chosen activity such as completing workbook assignments. Meanwhile, the teacher should continue to work with the rest of the group until all children master the objectives, perhaps dismissing them one by one as they do.

The teacher should be careful to avoid negative statements regarding the children who remain for extra help. The children who have mastered the lesson should be dismissed without fanfare and without calling attention to the fact that they have succeeded. Similarly, the remaining children should not get the impression that they have failed or done something wrong because they remain in the group. One way to handle this situation might be to let the group know in advance, perhaps in the introduction, that they might have different activities to complete after reading group. When the time comes to split up the group, the teacher could dismiss those who have mastered the objectives with instructions to complete a workbook activity or go to a learning center, for example. She could then continue with the remaining children, either with the original lesson plans or with another activity such as those suggested in the teacher's manual for children with individual needs. If the children ask why they are staying behind, the teacher should answer positively ("Everyone has different work to do. This is your work for today.").

15. Sometimes the teacher may wish to use one or more children who have mastered the objectives to serve as models for the others. This may be done with the group intact, or the teacher may dismiss all but the models and the children who need extra help. Sometimes, children having difficulty attaining objectives may benefit more from observing interactions between the teacher and students who already understand the process than they would from being questioned themselves. For example, the teacher might be teaching the difference between the sounds of words like tape and tap (to present the idea of a final e making the vowel sound long). The teacher might ask a child who does not understand this concept to read pairs of words, then give him the answers each time with an explanation. But doing this repeatedly for several pairs of words may prove frustrating for both child and teacher. Thus, it may make more sense for the teacher to keep children who have mastered the objectives in the group in order to carry on a dialogue or demonstration with them and provide a model for the other children. She can then turn her attention back to the others after they have had additional opportunities to see and hear the answer and explanation modeled several times.

An advantage of using models is that children often pay attention to and imitate peers whom they respect and like. Therefore, the teacher should be careful to help the model maintain the respect and friendship of his peers, and to prevent any resentment which might arise if the situation is mishandled. In particular, the teacher should not make the other children look bad while making a good example of the model(s). Instead of saying something like, "Janet's so smart. She knows the rule," the teacher should contain her comments to the answer itself, not the students' abilities. When commenting on a model's answer, the teacher should be specific about what was correct and why, since this helps the other students to focus on the important aspects of the problem ("That's right, you looked at that last letter to see if it was an e before you said the word.")

16. If one or more children still do not succeed in meeting the objectives within the time available for the lesson, provision should be made for tutorial assistance. This might come from the teacher herself or from peers who have mastered the lesson's objectives. In any case, students who fail to meet objectives should receive extra help, and must not be allowed to fall progressively behind. As stated in principle 14, such assistance should be given in a positive manner so that the children do not get the impression that they have failed or done something wrong. The suggestions in the teacher's manuals can, of course, be used for activities. When the teacher's time is limited, parent volunteers or older students might serve to help these students;

PART II: RESPONDING TO CHILDREN'S ANSWERS

The previous section dealt with group management practices. The second part of the system is concerned directly with the teacher's role in dealing with individual students within the group. The teacher has two major responsibilities in an academic exchange with the child: she must present the question, and then respond to the child's answer with feedback of some sort. The following principles focus primarily on the latter.

Three distinctions will now be considered in turn, and then will be used to explain how the teacher decides what kind of feedback she will give. These distinctions are types of question, types of learning and appropriate pacing, and types of child answers.

Questions

There are two basic types of questions. The first is a question that calls for a short, factual answer. These often deal with matters of fact which one either knows or does not know. Answering such questions requires remembering information. Thus, it is not possible to "think them out." Questions of this sort usually start with who, when, what, and where and might entail supplying labels or dates, or reading sight words. For example, when asked "What shape is this?" a child either does or does not recall the name. Generally, he cannot be helped with a clue.

The second kind of question can be reasoned out. This includes some who, what, when, and where questions that ask for more than a label (such as a question about story content.) This type of question also includes how and why questions which do not have short factual answers. Examples are, "Why do Eskimos wear warm clothes?" and "How can you tell when it is time to get up?" Giving the children clues can help them to reason or remember answers to these questions. For example, a clue to the Eskimo question might be, "Would you wear a bathing suit in the snow?" Then after a response, "Why not?"

These two different types of questions make different demands on the child. Purely factual or labeling questions call on memory alone, while other questions also may call on reasoning processes.

Types of Learning and Appropriate Pacing

Different types of learning will require different strategies in pacing the speed of questioning. The distinction to be made here is between 1). demands for rote memory suitable to drill and 2). questions requiring reasoning which cannot be answered automatically. Examples of rote learning are recognition of sight words and recitation of the multiplication tables. The children are expected to respond quickly to such questions without having to stop and think. Reasoning demands ask a child to apply a process (such as a word attack skill) or give an answer which requires some thought, such as memory of story content or an opinion. In general, rote learning is more easily accomplished with a rapid pace, while demands requiring more thinking should be presented with a slower pace. The teacher must decide what demands she is making of her children, and then set the pace which will best meet the objective.

In a rapidly paced lesson, the teacher moves quickly from child to child. The purpose of such a pace is to provide each answer many times, so that the children can learn through rapid repetition to recognize words, letters, etc. on sight "automatically." The child learns to do this from hearing and seeing repeatedly the association between the question and answer.

304

A rapid pace can be maintained when short feedback is given rather than elaborate feedback. The teacher waits only two to three seconds for a child to respond. If there is not a response, the answer is given and the teacher moves on. Appropriate feedback during rapid pacing is further discussed in principles 17, 18, and 19.

A slower paced lesson is one in which the teacher spends more time with each child and each question and gives more extensive feedback. This type of pace is suitable for demands requiring reasoning or use of new skills. In these situations the child learns by doing the process or by seeing it done and explained. For example, learning to sound out new words with certain combinations of sounds is a more complex process than the simple associative learning of common words as described above. This second type of learning often requires explanations and the process of getting an answer is usually more important than the answer itself. The feedback to be used in a slower-paced lesson is also discussed in more detail in principles 17, 18, and 19.

Type of child answers

Children's answers may be classified as (1) mostly or all correct, (2) mostly or all incorrect (we include in this category the answer "I don't know," which indicates a lack of knowledge), or (3) no response at all. Each of these situations requires a different response from the teacher, depending on the demands of the question and the capability of the child.

The rest of the principles are based on the premise that any child's response can be turned into a pleasant learning experience by the teacher. Therefore, wrong answers and "I don't know" statements are not undesirable in themselves. They can be used to promote learning when handled well.

However, a failure to respond is not desirable, and the child should be encouraged to respond in some way, even if to say "I don't know." It is then the responsibility of the teacher to leave the child with a good feeling about having responded, even if it was only to listen to the correct answer and repeat it.

The rest of the principles discuss teacher feedback to different types of child answers. Types of questions and types of pacing are discussed under each category. A summary of appropriate use of feedback appears in chart form following the discussion of principles 17, 18, 19, and 20.

WHEN THE CHILD DOES NOT RESPOND

17. When the teacher asks a question or gives a direction, she should wait for the child to respond, and also see that the rest of the group waits and does not call out answers. The length of time spent waiting for an answer depends on what kind of pace the teacher wishes to keep up. When the group is moving through rapidly paced questioning (such as drilling on sight words), she should wait only a few seconds and proceed by giving the answer herself if the child does not respond. However, when the pace is slower, the teacher should continue to wait for a response for as long as the child looks like he is thinking about an answer and may come up with one. However, she should not wait so long that the group's attention is lost or the situation becomes anxiety-producing for the child. The teacher must decide on the spot what is the optimal wait-time.

If the teacher is unsure about whether a child is still thinking about a problem or whether he is completely stumped, she should ask him ("Do you know?" "Can you do it?") and then proceed on the basis of the child's response.

If the child says he does not know or cannot do it, the teacher should refer to the discussion of "When the child is wrong or does not know" (Principle 17).

If the child still does not respond within a reasonable time during slower paced questioning, the teacher should provide help, by simplifying the question by degrees but always attempting to get some response (see strategies for simplifying questions below). If necessary, she should at least get a "yes-no" answer to the question "Do you know the answer?" By making sure to get an overt verbal response to every question she asks, she will gradually condition the children to respond to every question.

If the child still doesn't respond, or if he finally responds incorrectly or says "I don't know," the teacher should follow procedures for simplifying as discussed below, in #19.

A child who persistently requires encouragement to respond will probably require some tutorial assistance and should probably receive fewer and easier group demands until he is more willing and able to respond.

WHEN THE CHILD IS INCORRECT OR DOES NOT KNOW THE ANSWER

18. Wrong answers and "I don't know" statements should not be met with negative reactions by the teacher. If the child responds incorrectly, the teacher should first tell him that the answer is not right. She can do this by using such phrases as "No, that's not right" delivered in a non-critical voice, or she can acknowledge that the answer was partly right or that the child was using the right process but misapplied it or didn't complete it: "That's good; you remembered to think about those beginning and ending sounds, but the word isn't right--it doesn't make sense there." In pointing out that the answer is wrong, the teacher should be as specific as possible

about what was wrong.

When the child has answered incorrectly, the teacher should follow the simplification procedures outlined in the next section. These are the same procedures to be used with a non-responsive child, usually after the teacher has elicited at least a statement of "I don't know" from him.

SIMPLIFICATION PROCEDURES

19. After attempting to get a response or acknowledging that an answer is wrong, the teacher has two options for simplifying the question. She may give the correct answer to the child, or she may rephrase the question and give clues.

a. If the question deals with a matter of fact, so that the response cannot be reasoned out, the teacher must give the child the answer. She should not ask another child to provide it. Calling on others in this way can create bad feelings and over-competitiveness. Also, this may convince the children that if they do not respond or don't try to answer correctly, the teacher will eventually go on to someone else. Staying with the child until an answer is established and attempting to elicit some acceptable response from him will instead teach the children that they must listen, think, and respond.

Providing the answer to the child can be done in several ways. If the pace is rapid, the teacher should give the answer and move on, perhaps occasionally having a child repeat the response. If the pace is slow, the question can be restated in a form which simply calls for agreement, repetition, or choosing between alternatives. For example, the question "What punctuation mark is this?" can be simplified to "Is it a comma or a period?" Here, the child only has to make a choice. If the choice is still too difficult, one of the options can be made more apparent, such as "Is it a comma or a question mark?" An extension of this is to make one of the alternatives so ridiculous that the child not only enjoys it but sees that the correct choice is obvious ("Is it a comma or a worm?"). The child might also be given the opportunity to make a yes-no choice by questions such as "Is it a comma?"

Another strategy for simplifying factual questions which leads almost certainly to a correct response is to give the answer and ask the child to repeat it. For some children, repetition may be the only demand to which they can comfortably respond at first.

Giving the answer to the child in the form of a simplified question to which he can respond enables the child to succeed. This is particularly important for children who are anxious about responding or who seldom get a right answer. With children who generally reply quickly and correctly, it usually is not necessary to always provide a success experience, especially when the questioning is rapidly paced and the teacher knows that the child will not react negatively to being told the right answer.

b. If the question is such that the child can be expected to figure it out if given help, the teacher should give clues or rephrase the question in a way that guides the child's thinking in the right direction. If the clues do not help and giving the answer is necessary, the teacher should give it herself rather than call on another child for it.

One way to rephrase a question might be to break it down into a sequence of related questions. For example, in reviewing a story from the day before, the teacher might ask, "How did Tom make the bread?" If the child could not remember all the steps, the teacher could break it down into, "What did he do right after he decided that he wanted to make it?" Then, after an answer, "What did he do after he got home from the grocery store?", etc.

Another example of breaking a question down into sequential steps might be in helping a child sound out a new word. ("What is the beginning sound?" [response], "What do you know about those vowels in the middle?" [response], "Read the rest of the sentence. What word makes sense there that has the sounds you just read?")

Another way to rephrase a question, especially a "why" question, is to help the child focus on relevant aspects of the situation. For example, if the question was, "Why is there a railing around the Tiger Pit?", the teacher might say, "Well, if there was no railing around the cage, what might happen if the tiger decided to take a walk?". Or she might ask, "What would happen if the tiger got out of his cage?", and then, after answers, "When the rail is there, can the tiger get out of his cage?"

Sometimes, rephrasing of the question into simpler language may be sufficient to help the student. For example, "Name me some reasons that Tom and his family were eager to get started," might be more easily understood as "Why did Tom's family want to go?"

Simplification, therefore, involves breaking a question down into a simpler form that helps the child direct his thinking to the right answer. With questions requiring reasoning or application of a skill, the methods of simplification can be more complex and extensive than the methods described above (under heading 19 a) for factual questions.

If simplification of reasoning questions does not help the child get on the right track, the teacher should supply the answer, along with an explanation of the thought process involved in figuring it out ("You have to say the beginning sound, then the end sound, then look to see if you know anything about the letters in the middle. Then think what word has those sounds and makes sense there.") Again, in supplying the answer to a child, the teacher should try to finish with a question or response demand that the child can handle successfully, especially if the child is shy or apprehensive about responding.

WHEN THE CHILD IS CORRECT

20. If the child answers correctly, the teacher should acknowledge it. This can be done briefly by a nod, by repeating the answer, or by verbal indication of agreement, such as "right", "okay", etc. Praise may or may not be appropriate, and is further discussed in principles 21 and 22 below. After acknowledging a correct answer, the teacher should make sure that the rest of the group has heard and understood. If the others did not hear, she should have the child repeat the answer more loudly. The teacher might also repeat the answer herself and paraphrase it, although she should not get into the habit of following every answer with repetition, since the children may stop listening to one another's answers. Sometimes she should ask another child to repeat the answer, as discussed in principle 9.

Teacher Responses to Children's Answers: Summary

This chart summarizes principles 17, 18, 19, and 20. The teacher bases her choice of feedback on the type of question, the pace, and the child's answer.

TYPE OF QUESTION	PACE	CHILD'S RESPONSE	FEEDBACK RESPONSE
Factual (child cannot be helped to figure out the answer if he does not know)	Rapid	No response	Teacher (T.) waits only a few seconds, gives the answer and moves on.
		Incorrect or mostly incorrect	T. says that the answer is incorrect, gives the correct answer and moves on.
		Correct or mostly correct	T. acknowledges that the answer is correct. She makes sure all have heard the answer and moves on.
	Slow	No response	T. waits longer for a response (as long as the child seems to be thinking about it), then she simplifies it to get some overt verbal response, and deals with the verbalization as correct or incorrect.
		Incorrect or mostly incorrect	T. tells the student the answer is incorrect, simplifies the question by giving a choice or gives the answer and lets the student repeat it.
		Correct or mostly correct	T. acknowledges the correctness and makes sure all have heard.
Reasoning (child can be helped to figure out the answer).	Slow	No response	T. waits for a response as long as the child seems to be thinking about it, simplifies the question to get some verbal response, then deals with the verbalization as correct or incorrect.
		Incorrect or mostly incorrect	T. tells the student the answer is incorrect or partly correct, comments on the process where appropriate, then simplifies the question.
		Correct or mostly correct	T. acknowledges that the response is correct, comments on the process where it is appropriate, and makes sure all have heard.

PRAISE AND CRITICISM

21. Praise is an important aspect of teaching and should be used regularly but not indiscriminately. When used sincerely, it can reinforce desired behaviors and favorably influence the childrens' attitudes about themselves and school. The teacher should take care to praise the child's effort and/or thinking processes used in arriving at an answer, not just the answer itself. The teacher should use a variety of praise statements rather than rely on a single stock phrase, and should avoid praising too frequently lest her praise become taken for granted. During rapidly paced portions of the lesson it is probably best to avoid praise altogether, and instead confine responses to confirming students' answers and repeating the correct answers. When the pace is slower and students are called upon to demonstrate newly learned or more difficult skills, the teacher should begin praising more frequently.

Praise should be specific and individualized for each student. By being specific in her praise, the teacher can help the student focus on appropriate behaviors to be repeated. (This also helps make other students aware of what aspect(s) of the response were correct.) By making praise contingent on individual progress, the teacher can help each child see and appreciate his own progress (rather than praising only behaviors which some children have mastered and othershave not). In other words, individual progress rather than group norms should be the basis for praise of individual students.

For example, a child who usually gives up easily on new words but who, one day, does sound out a word should be told, "Mary, that was good; you looked at the word and then sounded it out by yourself." However, a child who consistently sounds out new words but needs to work on pausing at the end of sentences could be praised for doing the latter: "Good, John, I liked the way you waited after the

313

periods." In these examples, both Mary and John were praised for specific behaviors which indicated progress for them. By specifying the behaviors in these ways, the teacher gave more information to the children than if she had just said, "Good, Mary" or "Good, John."

22. Children who blurt out answers, call out answers out of turn, respond impulsively, or continue to respond the same way time after time regardless of the question should be corrected, but correction should come in the form of criticism combined with specific positive instructions about what was wrong and what should have been done. ("Don't just guess, think about the problem first before you try to answer." or "Don't pick out another activity now. The bell has just rung and remember, that means it's time for you to go to reading group.") Criticism along, without the additional provision of positive, prescriptive information about what to do instead, will be of little use to the child and may be harmful if it makes him inhibited or resentful.

Appendix B

Summary of the Observation System Used in the First-Grade Reading Group Study

This appendix summarizes the system used to collect the data described in the report. Full instructions about using the system are contained in the manual (Brophy, et al., Note 4).

Figures 1 and 2 present the forms used with the system, and they are numbered to correspond to the descriptions given below.

Appendix C contains a glossary in which terms used to describe the data are listed in alphabetical order with definitions.

The observation system is composed of two parts: a section for recording group data (information about events affecting the group as a whole), and individual data (information about the teacher's interactions with individual students).

Group Data

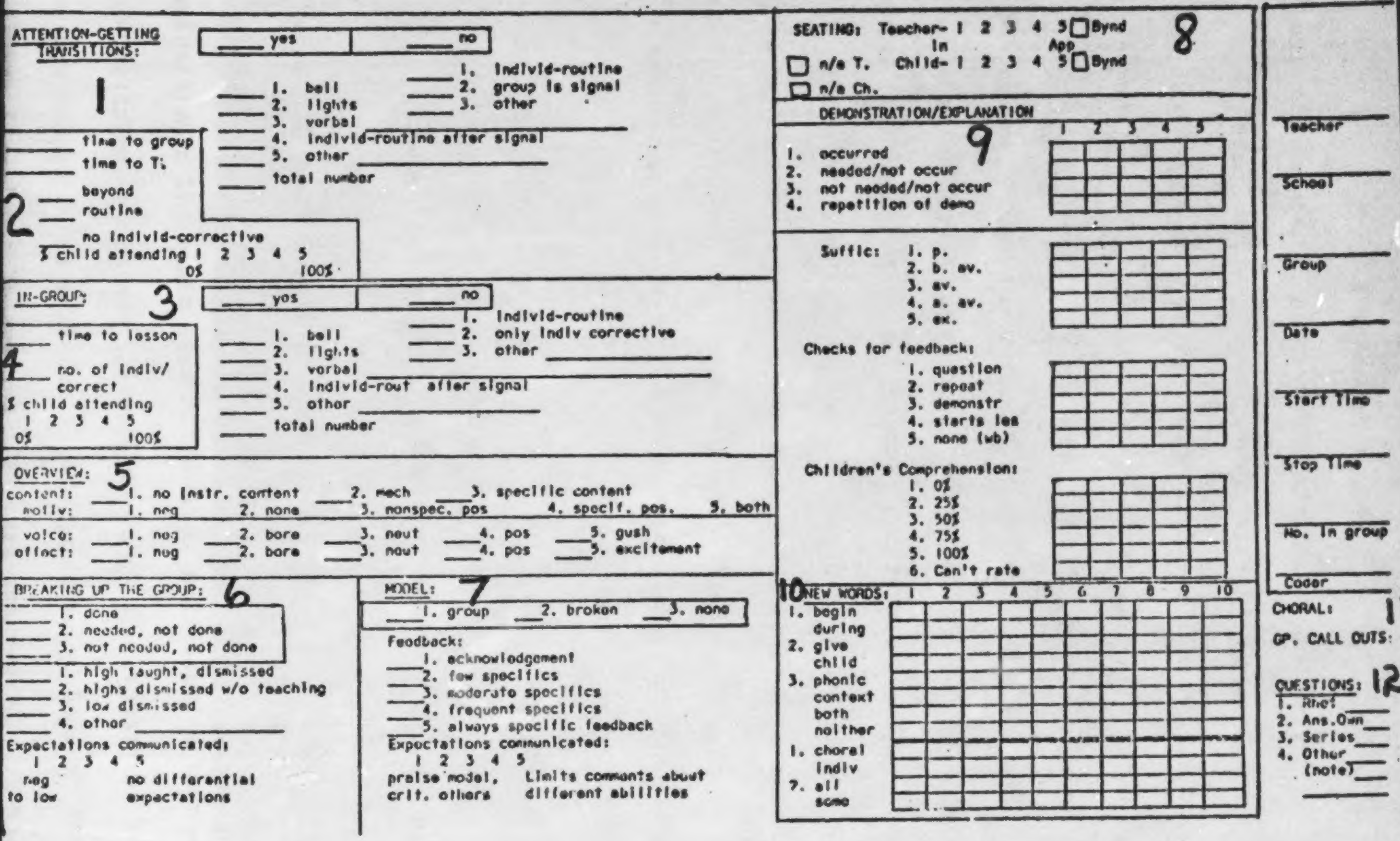
Figure 1 presents the form used to record group data. One such sheet is completed for each observation of a reading group.

1. Attention-getting transitions. The observer checks either "yes" or "no" to indicate whether or not the teacher has used a general attention-getter (that is, some signal that is delivered to the entire class to tell them that a transition is about to take place and a particular reading group is about to start). If "yes" is checked, the coder indicates which kind of signal is used by checking one of the five types listed underneath the "yes" category. The options are: 1) the use of a bell, 2) turning the lights

Figure 1

First Grade Reading Group Study

Group Coding Sheet



ATTENTION-GETTING TRANSITIONS: 1

____ yes ____ no

____ 1. bell
____ 2. lights
____ 3. verbal
____ 4. individ-routine after signal
____ 5. other ____
____ total number

____ 1. individ-routine
____ 2. group is signal
____ 3. other

____ time to group
____ time to T.
2 ____ beyond
____ routine
____ no individ-corrective
% child attending 1 2 3 4 5
0% 100%

IN-GROUP: 3

____ yes ____ no

____ time to lesson
4 ____ no. of indiv/correct
% child attending
1 2 3 4 5
0% 100%

____ 1. bell
____ 2. lights
____ 3. verbal ____
____ 4. individ-rout after signal
____ 5. other ____
____ total number

____ 1. individ-routine
____ 2. only indiv corrective
____ 3. other ____

OVERVIEW: 5

content: ____ 1. no instr. content ____ 2. mech ____ 3. specific content
motiv: 1. neg 2. none 3. nonspec. pos 4. specif. pos. 5. both
voice: ____ 1. neg ____ 2. bore ____ 3. neut ____ 4. pos ____ 5. gush
affect: ____ 1. neg ____ 2. bore ____ 3. neut ____ 4. pos ____ 5. excitement

BREAKING UP THE GROUP: 6

____ 1. done
____ 2. needed, not done
____ 3. not needed, not done
____ 1. high taught, dismissed
____ 2. highs dismissed w/o teaching
____ 3. low dismissed
____ 4. other ____

Expectations communicated:
1 2 3 4 5
neg to low — no differential expectations

MODEL: 7

____ 1. group ____ 2. broken ____ 3. none

Feedback:
____ 1. acknowledgement
____ 2. few specifics
____ 3. moderate specifics
____ 4. frequent specifics
____ 5. always specific feedback

Expectations communicated:
1 2 3 4 5
praise model, crit. others — Limits comments about different abilities

SEATING: 8

Teacher- 1 2 3 4 5 ☐ Bynd
In App
☐ n/a T. Child- 1 2 3 4 5 ☐ Bynd
☐ n/a Ch.

DEMONSTRATION/EXPLANATION 9

1 2 3 4 5

1. occurred
2. needed/not occur
3. not needed/not occur
4. repetition of demo

Suffic:
1. p.
2. b. av.
3. av.
4. a. av.
5. ex.

Checks for feedback:
1. question
2. repeat
3. demonstr
4. starts les
5. none (wb)

Children's Comprehension:
1. 0%
2. 25%
3. 50%
4. 75%
5. 100%
6. Can't rate

10 **NEW WORDS:** 1 2 3 4 5 6 7 8 9 10

1. begin during
2. give child
3. phonic context both neither
1. choral indiv
7. all some

Teacher ____
School ____
Group ____
Date ____
Start Time ____
Stop Time ____
No. in group ____
Coder ____

CHORAL: 11

GP. CALL OUTS:

QUESTIONS: 12
1. Who? ____
2. Ans. Own ____
3. Series ____
4. Other (note) ____

Figure 2

First Grade Reading Group Study

Individual Coding Sheet

Teacher______________________ Group__________ Time start_______ Page __ of __

School ______________________ Date __________ Time stop________ Coder_______

No. in group__________ ____ of_______

CONTEXT: 1
___ 1.slow, no WB
___ 2.WB
___ 3.Drill
___ 4.New reading
___ 5.Rereading

2
Teacher out:
1.________Byd____
2.________By____
3.________Byd____

Child behaviors: 15
1.Callout unaccepted
2.Callout accepted
3.Indiv inapp
4.social
5.private disturb
6.posture
7.materials
8.contraband
9.other in-group
10.praiseworthy
OUT-OF-GROUP
11.C init brief
12.C init long
13.T init brief
14.T init long

RESPONSE OPPORTUNITIES

		SELECTION										QUESTIONS									ANSWERS				FEEDBACK													RESULTS		BEHAVIORAL CONTACTS									
		Initial					Subs					Read				Non-read															Term			Sust															
3		4							5	6		7									8				9	10					11			12				13		14		15			16				
CHD# M	F	ORDRK	PRE	NVOL	VOL	CALL	CONT	ERROR	COMM	TURN	END	REP	RdCNC	WdREC	WdATT	PERS	CHOIC	PROD	COMP	INTRP	CORR	INC	DK	NR	NoFB	EMPH	FRCS	PRAIS	CRIT	SPEC	GA	AO	CALL	NEWQ	RPT	CLUE	GIVE	IMP	NoIMP	CHD# M	F	BEH#	PRAIS	IGNOR	MGMT	WARN	CRIT	EXIT	SPEC

on and off, 3) a verbal signal, 4) the delivery of a signal and then the routine contacting of all individuals involved, and 5) any other general signal.

The total number of general signals is entered in the blank which is titled "total number." If no general attention-getter was used by the teacher for the entire group, this is indicated by checking "no," and one of the three categories under "no" is checked. Those categories are: 1) "individual-routine," in which the teacher contacts each child who is supposed to come to reading individually, 2) "group is signal," in which the children respond to the dismissal of the previous group and come to the reading group without being reminded by the teacher, and 3) "other" which is used to designate any other way that the teacher might cause a transition without delivering a general signal to the entire class.

2. Further descriptions of transitions. "Time to group" is the time in minutes (to the nearest half-minute) that elapsed between the first signal that called a group to the circle and the arrival of the last student in the reading group area. "Time to T" is the time elapsed in minutes (to the nearest half-minute) between the arrival of the last student in the reading group and the arrival of the teacher. Also in this section, the observer notes whether any time noted for "time to T" was beyond her control or was due to routine managerial problems. The "number of individual corrective contacts" is entered as the total frequency of children contacted by the teacher because they did not attend to the transition signal properly. The "percent of children attending" to the signal is determined by the coder and is entered on the scale of 1 to 5, where 1 indicates that none of the children attended to the signal immediately, and 5 indicates that all of the children attended immediately.

3. Ιn-group attention-getting. Again the coder indicates by checking either "yes" or "no" whether or not the teacher used a signal delivered to the entire group in order to get their attention at the beginning of the lesson. If "yes" is checked, then one of the five items underneath the "yes" heading is checked to indicate what kind of signal is delivered. The categories here correspond to the categories described above under "attention getting for transitions." If no signal is given then one of the three categories under the "no" heading is checked. These are 1) "individual routine", in which the teacher contacts each child individually to get his or her attention, 2) "only individual corrective" contacts, in which the teacher does not call for anyone's attention unless the child has misbehaved or is not attending, or does not call on anyone because no corrective contacts were needed, and 3) "other" which is checked and completed if the teacher uses any other way of calling for a child's attention rather than addressing a signal to the entire group.

4. Further descriptions of in-group attention-getting. "Time to lesson" is the time in minutes (to the nearest half-minute) that elapsed between the teacher's arrival in the group and the beginning of the lesson (defined as the first academic content addressed to the group as a whole). The "number of individual corrective contacts" is a count of the children corrected because they were not paying attention or were not ready to begin the lesson. The "percent of children attending" at the beginning of the lesson is rated on a 5-point scale in which 1 indicates that no children were attending at the beginning of the lesson without correction and 5 indicates that all of the children were attending at the beginning.

5. Overview. The measures included here describe whether or not the teacher gave a general overview at the beginning of the lesson and, if so,

how it was delivered and what effect it had. The content of the overview is designated by checking either "no instructional content" to indicate that no such statement was given, "mechanical content" to indicate that only a descriptive overview was given, such as, "Today we will read pages 58-62," or "specific content" to indicate that the teacher gave an overview that explicitly described what lesson content would be covered that day. The motivational component of the overview is described by checking one of the five categories, either "negative" to indicate that the teacher made an overtly negative statement about how bad the lesson was going to be, "none" to indicate that the teacher made no statement regarding the motivational value of the lesson, "nonspecific positive statements" which indicate that the teacher said something like, "You are really going to enjoy what we are going to do today." "Specific positive motivational content" in which the teacher explained why the lesson was going to be good, such as saying "When we finish this lesson you will be able to play the word game that you have been wanting to play for so long," or "both," a category which includes both the nonspecific positive and the specific positive statement just described.

If either an overview regarding the content or the motivational effect of the lesson was given, then the "voice" and "effect" scales are completed. The "voice" scale indicates the tone of the teacher's voice and the apparent excitement or favorableness accorded by the teacher to the upcoming lesson. This ranges along a 5-point scale from _1_ indicating an obvious negative outlook to _5_ indicating gushing enthusiasm. The "effect" scale measures the children's reactions to the overview and ranges from _1_ indicating an overt negative reaction to _5_ indicating excitement on the part of the children.

6. _Breaking up the group_. This set of items describes whether or not the teacher broke up the group due to ability differences and, if so, how this

was handled. The coder checks either that the group was broken up by indicating "done," that the group should have been broken up because of ability differences, but was not by checking "needed, not done," or that the group did not need to be broken up and consequently was not by checking "not needed, not done." If the category of "done" was checked, then the following items in the section are also completed. The first four items listed describe how the group was broken up. "High taught, dismissed" indicates that the students who were grasping the lesson more quickly than the others were taught through to the end of the objective and then sent to their seats so that the teacher could work further with those students who were not learning as rapidly. "High dismissed without teaching" indicates that the teacher went no further with the high-ability students and simply dismissed them and retained the lower-ability students and worked with them for the rest of the lesson. "Low dismissed" indicates that the teacher sent away those children who were not catching on as quickly as the others and spent the rest of the group time working with those children who were learning rapidly. The category of "other" is checked and completed for any other method of breaking up the group due to ability differences. The "expectations communicated" by breaking up the group is measured on a 1 to 5 scale with 1 indicating overt negative statements to the lower ability students, in which the teacher communicated to everyone that certain students were not achieving as well as other students. A score of 5 on this scale indicates that the teacher managed to break up the group without overtly communicating any differential expectations.

7. *Use of a student model*. The coder checks one of three items to indiindicate whether or not the teacher used another student as a model for the rest of the group in an academic interchange. "Group" is checked if the

teacher did use a model while the group was intact, without sending away any of the students because of ability differences. "Broken" is checked if the teacher used a model after breaking up the group due to ability differences and retained one of the more capable students along with other students who needed some extra help and then used the more capable student in an interchange to model questions and answers for the rest of the group. "None" is checked if no use of the model is present. If a model is used in any way, then the rest of this section is completed. The items listed under the heading "feedback" describe how the teacher responded to the model's answers along a 5-point scale. A score of _1_ indicates mere acknowledgement of correctness and _5_ indicates that the teacher gave very specific feedback with regard to the model's answers so that the rest of the group could understand what was right about those answers, not simply that they were right. A scale of "expectations communicated" is also completed. A score of _1_ on this scale indicates much use of praise for the model with criticism of others in the group and direct comparisons of the ability of the model to the ability of the other children. A score of _5_ indicates that the teacher made no comments about differential ability and focused entirely on the answers given and the content of the lesson.

8. _Appropriateness of seating_. Appropriate seating is defined as: 1) the teacher being placed so that he or she may see the rest of the children in the room as well as the children in the reading group and 2) the placement of the children so that they can see only the teacher and their vision of the rest of the room is hampered as much as possible. Two scales are therefore completed for the appropriateness of seating, one for the teacher and one for the child. In both cases _1_ responds to very inappropriate seating and _5_ responds to appropriateness. The points in

the scale are determined by the percentage of error. For example, 3 on the scale for the teacher would indicate that the teacher could only see 50% of rest of the class. A point of 3 on the children's scale would indicate that 50% of the children in the group could see what was going on in the rest of the classroom. In either case, if the seating arrangement is inappropriate because of building restraints or some other factor beyond control of the teacher, then the "beyond" box is also checked. If for some reason this scale is not applicable for either the teacher or the children, then "N/A Teacher" or "N/A Children" is checked. The item is not applicable for the teacher whenever there is another adult in the room so that the teacher does not have to supervise the rest of the children. The item is not applicable for the children when there are no other children in the room and nothing else which would distract the children so that it does not matter how they are seated.

9. Demonstrations/Expectations. The items in this category are completed for each activity assigned by the teacher, up to five activities. These usually involve workbook or seatwork assignments, which must be explained to the children so that they know what to do. The numbers at the top of the grids refer to the activity (first, second, etc.) and for each one, the vertical column is completed. The coder first determines for each activity whether a demonstration has "occurred," was "needed, but did not occur," was "not needed and did not occur" even though an activity is about to take place, or whether there is a "repetition of a demonstration" because a previous one was insufficient. (The coder determines whether or not a demonstration or explanation is needed by the reaction of the children.) If a demonstration has occurred, then the coder completes the "sufficiency" scale, which ranges from a score of 1 indicating a poor demonstration or explanation to a score

of _5_ indicating an excellent demonstration or explanation. The sufficiency is determined by the coder after considering the amount or detail that was included, the sequencing of the steps explained, and the overall clarity of the explanation. The coder also indicates whether any checks for feedback were made by the teacher and what kind were made. "Question" refers to a call for questions. (The teacher says something to the effect of, "Does everyone understand?" or "Are there any questions?") "Repeat" indicates that the teacher has asked a child or children to repeat the directions. "Demonstrate" means that the teacher has asked a child or children to demonstrate the activity by going through some examples. "Start lesson" is checked whenever the teacher is going to observe the children working on the activity and can therefore check on their understanding while they are doing this. "None (WB)" is checked whenever the teacher gives an explanation of a workbook activity, then dismisses the children to perform it, and does not have any checks for feedback. The "Children's comprehension" of instructions for activities is noted for each demonstration or explanation by a 5-point scale in which _1_ indicates that none of the children understood the directions and _5_ indicates that all of the children understood the directions. If the coder cannot rate children's comprehension because they were dismissed from the group before the activity was performed, the category of "Can't rate" is checked.

10. New words. New words are vocabulary words or letters which are introduced to the children and are used in the lesson in some way. They do not include words or letters which are the focus of an entire lesson. Five variables are considered for each new word or letter that is given, up to ten words or letters. For each new word, the coder indicates whether it was given at the beginning of the lesson or during the lesson when it first

appeared. The coder also indicates whether the teacher gave the word to the children or had the child figure it out. The third set of variables indicates whether the teacher gave phonics clues, context clues, or neither in presenting the new word. The last two sets of items are concerned with whether the teacher has the children repeat a new word or letter. If it is repeated, the coder indicates whether it was done by choral or individual repetition, and whether or not all of the students had the chance to repeat the word or whether only some of them did.

11. Choral and call out responses. The coder keeps a tally of the number of choral responses and group call outs occurring within that observation period. Choral responses are defined as total group responses which are asked for by the teacher with the expectation that all of the children in the group will respond with the same answer. Group call outs are defined as several children calling out the answer to a question, whether or not the teacher has designated an individual to answer the question. They differ from choral responses in that the teacher does not indicate beforehand that she wants more than one child to answer.

12. Undesirable question. "Rhetorical" is checked whenever the teacher asks a very obvious question and does not really expect an answer, such as "Isn't that a beautiful picture?" "Answers own" is checked whenever the teacher asks a question and answers it without calling on a child or waiting for a child to answer it. "Series" is checked whenever the teacher asks more than one question in a row without waiting for a child to answer. Any other "undesirable" questions are noted under "other" and described.

Individual Data

Figure 2 presents the sheets used to record academic and behavioral interactions. The number of these sheets used for each observation depends on the length of the reading group.

1. Context. The coder notes on each sheet the lesson context for all instructions on that page.

"Slow, No WB" indicates a slow-paced question and answer session in which no workbook was involved. That is, the teacher is asking questions, children are answering, and the pace is not fast enough to consider it a drill.

"WB" stands for workbook activities. In this case each individual child is making some response to each question whether or not this is indicated by an oral response which is coded.

"Drill" is checked whenever there is rapid questioning and answering by the children. The determining characteristic of this context is its rapidity. The teacher moves quickly from child to child and gives very little substantive feedback.

"New reading" is checked whenever the children are reading a story from their textbook for the first time. They have not seen the material before.

"Rereading" is checked whenever it is apparent to the coder, usually because the teacher has said something to that effect, that the children are rereading material that they have seen before.

2. Teacher Out. In these blanks the coder should note the time that the teacher was out of the group and also indicate by checking "beyond" if the teacher's absence on that particular occasion was due to some factor beyond his or her control, such as a child getting sick, a phone call from the office, etc.

The rest of this coding sheet is divided into two sections: "Response Opportunities" and "Behavioral Contacts." In the response opportunities section, each line corresponds to an interaction between the teacher and a child in which a question was asked, the answer was given, and the teacher responded to the answer in some way.

3. Child number. A child number is entered on each line to indicate which child in the group is interacting with the teacher. The number is entered under either "male" or "female" to facilitate later data analyses.

4. Type of selection. Each line also includes a check in one and only one of the seven categories listed under Selection. There are two major types of selections, either initial selection or subsequent selection. Initial selection is used the first time the child is called on in a series of interactions, and one of the two subsequent selection categories is checked for each continued interaction without interruption. The five types of initial selection are "order," in which the child is selected on the basis of the seating pattern, "preselect," in which the child is not selected on the basis of the seating pattern but is named before the question is asked, "nonvolunteer," in which the child is not selected on the basis of the seating pattern and is not called upon until after the question has been asked, "volunteer," in which the child is called upon because he or she has raised a hand and offered to answer the question, and "call out," in which a child has called out an answer to a question without being designated by the teacher to answer that question and the answer is accepted by the teacher. The two types of subsequent selection are "continue" in which the interaction with the teacher is continued for some other purposes other than correcting a previous error and "error correction," which is checked whenever the interaction is being sustained with a child for the purpose of correcting a previous error.

5. Comment. This category is checked if the teacher has called on the child for the purpose of commenting about another child's answer for any other

reason than giving the correct answer to an incorrect question.

6. Designation of reading turn. Either "turn" or "end of turn" is checked whenever an interaction has occurred within a reading turn. These categories are necessary for later data analyses to separate out single response opportunities and those which occurred within a reading turn, since the only lines of coding entered for a reading turn are those which involve an interaction with the teacher. That is, a child may read correctly all the way through a turn and have no corrections and interactions with the teacher, in which case only one line of coding with "end of turn" checked would be entered to indicate that the child did read. However, if a child read the same materal and had several mistakes and several interactions with the teacher because of those mistakes, there would be several lines of coding with "turn" checked. Therefore, each interaction occurring within a reading turn is designated by checking one of these two categories.

7. Types of questions. For each line of coding one of the nine categories listed under Questions is checked. Questions may be either reading questions or nonreading questions. The categories considered under reading questions are those which have to do with a child actually looking at a written symbol and saying a word or letter. Nonreading questions have to do with any other type of question that is asked during the reading turn that does not emphasize actually reading a written symbol. The categories of reading questions are "repeat," in which a child is asked to look at a word and repeat what the teacher just said; "reading choice," in which the child is asked to choose between four or fewer alternatives in a question which emphasizes reading to make the correct choice; "word recognition," in which the task involves looking at a word and calling its name; and "word attack," in which the question requires the child to analyze a word

or letter into its component letters or sounds or to answer a question about word attack skills. Types of nonreading questions are "personal" questions, which require the child to say something about his personal background or experience; "choice" questions which ask the child to choose between four or fewer alternatives in a situation that does not involve the reading of words in order to answer the question; "product" questions which require the child to give some kind of factual answer which is not related to the comprehension of the story just immediately read; "comprehension" questions which require the child to answer a question about the content of material which has just been read; and "interpretative" questions which require the child to go beyond some material which he or she has just read or heard and to either make predictions, evaluations, or interpretations about the material, as in, "What do you think is going to happen to David now?"

8. Answers. For each line of coding one of the four categories of Answers must be checked. The question can be either "correct," "incorrect," a child may answer "I don't know," or the child may make "no response."

9. No feedback. If the teacher makes no response to the child's answer and therefore does not give the child any indication of correctness or incorrectness, then this category is checked. No other kind of feedback is checked when "no feedback" has occurred.

10. General feedback. The five categories described here may occur in conjunction with other types of feedback with the exception of "no feedback." "Emphasis" is checked whenever the teacher repeats the child's answer for the rest of the group or has the child or another child repeat that answer for the purpose of the other children hearing it. "Process feedback" is checked whenever the teacher goes beyond acknowledging or giving an answer and actually explains the answer or the process used in arriving at that answer, such as

explaining the process of sounding out a word. "Praise" and "Criticism" are checked whenever they occur. These categories apply to praise and criticism of a child's answers only. "Specific" is checked in conjunction with either praise or cirticism when the teacher has gone beyond a simple statement of good or bad and specifies what was good about the answer or what should have been done differently to avoid criticism of the answer.

11. Terminal feedback. The three categories described here are used whenever the teacher does something to end that child's response opportunity about that question. The teacher can either "give the answer," "ask other," or accept a "call out," in which another child shouts out an answer or a correction before the teacher can respond and the teacher allows this to stand as feedback in termination of that question.

12. Sustaining feedback. The four categories listed here are checked whenever the teacher does something which sustains an interaction with that child. "New question" is checked when the teacher asks the child a new question which does not serve as an error correction of a previous answer. The other three categories describe an attempt to have the child correct a previous error. "Repeat" is checked whenever the teacher simply repeats the question or prods the child without giving any substantive help, "clue" is checked whenever the teacher gives some help by rephrasing the question or giving clues but does not simplify the question so much that the answer is obvious, and "give" is checked whenever the teacher gives a clue or rephrases the question in such an obvious way that the child is almost insured success on his second attempt at answering.

13. Results. Whenever one of the last three types of sustaining feedback is checked ("repeat," "clue," or "give") the result of that error correction feedback is recorded by indicating whether the next answer was

improved or was not improved.

The next block of categories concerns Behavioral Contacts by the teacher. These are recorded independently of the Response Opportunities.

14. Child number. The child number of the recipient of the behavioral contact is recorded if the child was in the group at that time. If a behavioral contact involves a child out of the group, then the number 998 is entered to indicate an out-of-group contact.

15. Behavior number. One of the fourteen behavior numbers listed in the upper right-hand corner of the coding sheet is entered in this column for every line of coding. The first ten are only applied to students in the reading group:

1. "Call out unaccepted" by the teacher and corrected.
2. "Call out accepted" by the teacher for its academic content but also corrected for the call out behavior.
3. "Individual inappropriate" behavior, which subsumes inattentive type behaviors or anything that a child does individually which is not appropriate to the reading group setting and does not fall into one of the other categories.
4. "Social interaction" with another child in the reading group which is not desired.
5. "Private disturbance" is an individual's behavior which is loud enough or active enough to disturb the other children.
6. "Posture" includes any inappropriate behavior related to sitting in the group.
7. "Materials" includes any inappropriate behavior

involving the books, papers, pencils, etc. used in the reading group activities. This includes holding the book upside down, being on the wrong page, misuse of the marker, etc.

8. "Contraband" includes the possession of any materials not appropriate to the reading group such as toys, gum, or books for other activities.
9. "Other in-group" inappropriate behaviors not listed above.
10. "Praiseworthy" behavior occurring during the reading group. This is only followed by a teacher reaction of praise. The first nine items are all considered behaviors which result in a corrective contact.

The next four behaviors are concerned only with out-of-group contacts: that is, the teacher interacts with a child that is not seated in the reading group. These behaviors are:

11. "Child-initiated brief," in which the child approaches the teacher and the contact is of a very brief duration (less than five seconds).
12. "Child-initiated long," in which the child approaches the teacher and the contact lasts longer than five seconds.
13. "Teacher-initiated brief," in which the teacher interrupts the group to say something to those outside the group but the contact is brief (less than five seconds).
14. "Teacher-initiated long," in which the teacher interrupts the group and the contact lasts longer than five seconds.

16. Teacher reactions. The next seven categories are checked to indicate what kind of teacher reaction occurred within each of the behavioral contacts. "Praise" indicates that the teacher said something about how good a particular behavior was. "Ignore" is used only for behavior categories of "call out unaccepted" and the "child-initiated out-of-group" contacts. "Management" is checked whenever the teacher delivered a corrective contact in a calm, nonirratated tone of vioce. "Warning" indicates that the teacher showed irritation or slight anger in the voice and implied a threat to the child or some punishment if he or she does not comply. "Criticism" is checked whenever the teacher's voice was extreme and demonstrated anger or delivered some punishment to the child because of the behavior. "Nonverbal interaction" is checked whenever the teacher delivers a corrective contact without saying anything; that is, a gesture of some sort served to correct the child. "Specificity" is checked whenever any of the previous categories just described (with the exception of "Ignore") was specific as to either what behavior is being praised or as to what alternative behavior is desirable in the case of the corrective contact.

Appendix C

Glossary of Terms Used to Describe Data from the First-Grade Reading Group Study

The following is an alphabetical list of terms derived from the observation system used to describe teacher and student behaviors of interest in the study.

Academic interaction. This is also described as a response opportunity, an oral interchange in which the teacher asks an academic question of an individual student who answers in some way and may receive feedback from the teacher.

Ask other. A type of terminal feedback, in which the teacher asks a second child to answer a question that was not answered correctly by the first respondent.

Behavior contact. An interaction in which the teacher corrects the child for some misbehavior or praises some behavior that was good. These interactions are not academic in nature.

Call out. A type of selection in which a student answers a question without having been designated by the teacher (either by seating position or by being named). It is also a type of terminal feedback when a second student calls out an answer or gives some kind of feedback after the first student's answer was incorrect. This occurs without the teacher calling on the second student to request the answer.

Choice question. A type of question in which the student must select from four or fewer alternatives to answer a question that is not about a word or sound.

Clue feedback. A type of sustaining feedback following an incorrect answer or failure to respond. The teacher provides clues or simplified questions to help the same student get a better answer.

Comment. A second student is asked by the teacher to comment on the answer given by the first student. This is not done in order to provide the correct answer, but to let the second student repeat the answer or expand it.

Comprehension question. A type of question in which the student provides information about some material that has just been read in the group.

Correct answer. A type of answer that is acceptable to the teacher.

Criticism (in academic interactions). A type of feedback that includes a very negative response to the child and/or the answer.

Criticism (in behavior contacts). A type of teacher reaction to misbehavior that included an angry tone of voice and/or the delivery of punishment.

Don't know answer. A type of answer in which the student simply stated, "I don't know."

Emphasis feedback. A type of feedback by which the teacher ensures that the rest of the group heard the answer and understood it. This usually means that the teacher repeated it or had a child repeat it.

Feedback. The response offered by a teacher to an individual student's answer to an academic question.

Give answer feedback. A type of terminal feedback in which the teacher supplies the answer to a child who has not answered correctly.

Give by clue feedback. A type of sustaining feedback in which the teacher continues to ask questions to the child, but makes the new questions so simple that the answer is essentially given to the child. The difference between this category of feedback and that of give answer is that this technique elicits the correct answer from the child who originally made the error.

Ignore. A type of teacher reaction to misbehavior in which no acknowledgment is given by the teacher. This category was only applicable to call outs and child-initiated out-of-group interruptions.

Improvement. The result of sustaining feedback that led to a better answer than that originally given by the child.

Incorrect answer. A type of answer that was not acceptable.

Interpretation question. A type of question in which the student is required to make predictions, evaluations, or interpretations.

Management. A type of teacher reaction to misbehavior that was mild in tone with no threat of punishment.

New question. A type of feedback in which the teacher asks a related question of a student who has just completed answering a question.

No feedback. The absence of any feedback to an answer. The teacher did not respond with affirmation or negation of the answer and provided no other substantive feedback.

No improvement. The result of sustaining feedback that did not lead to a better response than that given by the child to the original question.

No response answer. A type of answer in which the child did not say anything in response to the teacher's question.

Nonreading questions. A large category of questions composed of five smaller categories: personal, choice, product, comprehension, and interpretation having in common that the student is being asked a question about a fact, the content of what has been read, or is asked for an opinion. See also reading question.

Nonturn response opportunities. The interaction between the teacher and child did not take place while the child was reading a sentence or passage aloud.

Nonverbal intervention. A type of teacher reaction to misbehavior that consisted of a gesture, look, or finger-snap.

Nonvolunteer. A type of selection in which the teacher asks a question and then calls on a child who has not volunteered to answer, and who was not previously designated to answer because of his or her seating position.

Ordered selection. A type of selection based on seating position because the teacher was using a pattern of selection, such as going from one end of the group to the other.

Personal question. A type of question in which the student was asked to describe a personal experience or state an opinion about personal preference.

Praise. A type of academic feedback or reaction to student behavior in which the teacher indicates a very favorable, positive response to the child, the answer, or the behavior.

Preselect. A type of selection in which the teacher names a particular child before asking a question and is apparently not selecting the student because of his or her seating position.

Process feedback. A type of feedback in which the teacher explains the steps or process necessary to solve a problem or answer a question.

Product question. A type of question in which the student must produce some fact or label from memory that does not include reading a word or answering a comprehension question.

Reading choice question. A type of question in which the student must choose between four or fewer alternatives that answer a question about a word, letter, or sound.

Reading question. A large category of questions composed of four smaller categories: repetition, reading choice, word attack, and word recognition, which have in common that the student is asked to decode or identify words, letters, or sounds. See also nonreading questions.

Repeat feedback. A type of sustaining feedback in which the teacher repeats the original question without providing any more information, or otherwise encourages the child to respond without providing any substantive help.

Repetition question. A type of question in which the student is asked to repeat a word that has just been read.

Response opportunity (R.O.). An interaction between a teacher and an individual child in the reading group about an academic question. It consists of three parts: the question asked by the teacher, the response of the child, and the feedback delivered by the teacher.

Result of sustaining feedback. Indication as to whether or not the teacher's efforts actually led to a better response by the child. This was classified as either improvement or no improvement.

Specificity. When used with praise, this term means that the praise specified exactly what was good about what the child just did or said. When used with criticism, it indicates that the teacher was specific about what should have been done instead of the criticized behavior.

Sustained interaction. A response opportunity (teacher question, child answer, and teacher feedback) which was the result of the teacher offering sustaining feedback to the incorrect answer just given by the child.

Sustaining feedback. Teacher response to a child who has answered incorrectly, has said "I don't know," or has failed to respond. The teacher stays with the child for another interaction instead of giving terminal feedback.

Terminal feedback. A response offered by the teacher when the child's answer was incorrect, "I don't know," or a failure to respond, ending the child's opportunity to answer that question by providing the answer or having it provided.

Total response opportunities. Response opportunities that are not distinguished as to whether or not they occurred during a reading turn.

Turn response opportunities. Academic interactions (response opportunities) that occurred while the child was reading a sentence or passage aloud.

Warning. A type of teacher reaction to misbehavior that included an irritated tone of voice and perhaps an implied or stated threat of punishment.

Word attack question. A type of question in which the student must give information about sounds of letters within a word, or about the process of breaking down a word into its component parts.

Word recognition question. A type of question in which the student must say the name of a word or letter by reading it aloud.